



**Universidad Nacional Mayor de San Marcos**

Universidad del Perú. Decana de América

Facultad de Ciencias Biológicas

Escuela Profesional de Genética y Biotecnología

**Análisis pangenómico de la bacteria patógena**

***Pasteurella multocida***

**TESIS**

Para optar el Título Profesional de Biólogo Genetista  
Biotecnólogo

**AUTOR**

Dennis Edgardo CARHUARICRA HUAMAN

**ASESORES**

Abelardo Lenin MATURRANO HERNÁNDEZ

Lima, Perú

2018



Reconocimiento - No Comercial - Compartir Igual - Sin restricciones adicionales

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Usted puede distribuir, remezclar, retocar, y crear a partir del documento original de modo no comercial, siempre y cuando se dé crédito al autor del documento y se licencien las nuevas creaciones bajo las mismas condiciones. No se permite aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros a hacer cualquier cosa que permita esta licencia.

## Referencia bibliográfica

---

Carhuaricra, D. (2018). *Análisis pangenómico de la bacteria patógena Pasteurella multocida*. [Tesis de pregrado, Universidad Nacional Mayor de San Marcos, Facultad de Ciencias Biológicas, Escuela Profesional de Genética y Biotecnología]. Repositorio institucional Cybertesis UNMSM.

---



**Universidad Nacional Mayor De San Marcos**  
(Universidad del Perú, Decana de América)

**FACULTAD DE CIENCIAS BIOLÓGICAS**

**ACTA DE SESIÓN PARA OPTAR AL TÍTULO PROFESIONAL DE  
BIÓLOGO GENETISTA BIOTECNÓLOGO  
(MODALIDAD: SUSTENTACIÓN DE TESIS)**

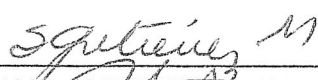
Siendo las 10:13 horas del 20 de diciembre de 2018, en el Salón de Grados de la Facultad de Ciencias Biológicas y en presencia del jurado formado por los profesores que suscriben, se dio inicio a la sesión para optar al Título Profesional de Biólogo Genetista Biotecnólogo de DENNIS EDGARDO CARHUARICRA HUAMAN.

Luego de dar lectura y conformidad al expediente N° 010-EPGB-2018, el titulando expuso su tesis: "ANÁLISIS PANGENÓMICO DE LA BACTERIA PATÓGENA *Pasteurella multocida*" y el Jurado efectuó las preguntas del caso calificando la exposición con la nota 20..., calificativo: Aprobado con mención honoris

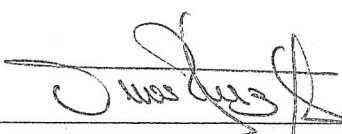
Finalmente, el expediente será enviado a la Escuela Profesional de Genética y Biotecnología, y al Consejo de Facultad para que se apruebe otorgar el Título Profesional de Biólogo Genetista Biotecnólogo a DENNIS EDGARDO CARHUARICRA HUAMAN y se eleve lo actuado al Rectorado para conferir el respectivo título, conforme a ley.


Siendo las 11:55 horas se levantó la sesión.

Ciudad Universitaria, 20 de diciembre de 2018.

  
Dra. SUSANA GUTIERREZ MORENO  
(PRESIDENTA)

  
Dr. LENIN MATURRANO HERNANDEZ  
(ASESOR)

  
Dr. DAN ERICK VIVAS RUIZ  
(MIEMBRO)

  
Blgo. MIGUEL NEIRA GONZALES  
(MIEMBRO)



***A mis padres por su esfuerzo y  
estímulo inquebrantables***

*"The most outstanding feature of life's history is a constant domination by bacteria"*

-----Stephen Jay Gould

*"Bacteria may not have sex often, but when they do, it can be really good, at least evolutionary"*

----Pål Jarle Johnsen

*"Though this be madness, yet there is method in't"*

----*Hamlet*, William Shakespeare

## **AGRADECIMIENTOS**

Al Dr. Lenin Maturrano, por su confianza y asesoría en el desarrollo del presente trabajo.

Al Dr. Raúl Rosadio por permitirme desempeñar y desarrollar esta tesis en el Laboratorio de Biología y Genética Molecular de la Facultad de Medicina Veterinaria, UNMSM.

A mis colegas del laboratorio Luis Luna, Juan Siuce, Jorge Maximiliano, Marjorie Aleman, Raquel Quino, Marcos Almeyda, Carla Durand, Grecia Pavia, Christian Changanahui, Laura Córdova y Susana Asencios por su compañía y amistad, y un reconocimiento especial a Raquel Hurtado y Natali Allasi.

A mis padres, Edgard Carhuaricra y Norma Huaman, sin su constancia nada sería posible. A mis hermanas, Tati y María, por su compañía y apoyo, un agradecimiento infinito.

Al Vicerrectorado de Investigación y Postgrado de la Universidad Nacional Mayor de San Marcos por financiar el presente trabajo con el Programa de Promoción de Tesis de Pregrado-2017 (proyecto A17080054b), gracias por su apoyo a la ciencia.

## ABREVIATURAS

MLST	:	<i>Multilocus Sequence Typing</i>
ST	:	<i>Sequence Typing</i>
CC	:	Complejo Clonal
LPS	:	Lipopolisacárido
PCR	:	Reacción en cadena de la Polimerasa
ERIC-PCR	:	<i>(Enterobacterial Repetitive Intergenic Consensus-PCR)</i>
REP-PCR	:	<i>(Repetitive Extragenic Polymorphism-PCR)</i>
CDS	:	Secuencia de ADN codificante
BLAST	:	Basic Local Alignment Sequence Tool
NCBI	:	National Center for Biotechnology Information
RPS-BLAST	:	(Reverse Position Specific -BLAST)
SNP	:	Polimorfismo de nucleótido único
RIRDC	:	Rural Industries Research and Development Corporation
OMPs	:	Proteína de membrana externa ( <i>Outer membrane protein</i> )
ADN	:	Ácido Desoxirribonucleico
RAPD	:	Amplificación aleatorio de ADN polimórfico
PFGE	:	Electroforesis en Gel de Campo Pulsado
ARNr	:	Ácido Ribonucleico ribosomal
COG	:	Cluster of Orthologous Genes
CDD	:	Conserved Domain Database
PSSMs	:	Position-specific score matrices
PMT	:	Toxina de <i>Pasteurella multocida</i>
RA	:	Rinitis atrófica
SH	:	Septicemia hemorrágica
CA	:	Cólera aviar
IS	:	Secuencias de Inserción

## ÍNDICE GENERAL

### LISTA DE TABLAS

### LISTA DE FIGURAS

### RESUMEN

### ABSTRACT

<b>1. INTRODUCCIÓN .....</b>	<b>1</b>
<b>2. MARCO TEÓRICO .....</b>	<b>2</b>
2.1. Enfermedades asociadas a <i>P. multocida</i> .....	2
2.1.1. Definición, impacto económico y amenaza zoonótica .....	2
2.1.2. Septicemia hemorrágica (SH) .....	3
2.1.3. Cólera aviar (CA) .....	4
2.1.4. Rinitis atrófica (RA) .....	5
2.1.5. Coriza .....	5
2.1.6. Pasteurelosis neumónica .....	6
2.1.7. Pasteurelosis como infección zoonótica .....	6
2.2. <i>Pasteurella multocida</i> .....	7
2.2.1. Historia y posición taxonómica .....	7
2.2.2. Identificación de <i>P. multocida</i> .....	7
2.2.2.1. Identificación fenotípica de <i>P. multocida</i> : caracterización microbiológica y bioquímica .....	7
2.2.2.2. Identificación genotípica de <i>Pasteurella multocida</i> .....	8
2.2.3. Tipificación de <i>Pasteurella multocida</i> .....	8
2.2.3.1. Tipificación serológica .....	8
2.2.3.2. Tipificación molecular .....	9
2.2.4. Factores de virulencia .....	10
2.2.4.1. Cápsula .....	10
2.2.4.2. LPS (lipopolisacáridos) .....	10
2.2.4.3. Proteínas de membrana externa (PMEs), fimbrias y adhesinas .....	10
2.2.4.4. Proteínas de metabolismo de hierro .....	11
2.2.4.5. Sialometabolismo .....	11
2.2.4.6. Toxinas .....	11
2.2.5. Patogénesis molecular y celular .....	12
2.3. Genómica bacteriana y de <i>Pasteurella multocida</i> .....	14
2.3.1. Secuenciamiento genómico en bacterias .....	14
2.3.2. Genoma de <i>Pasteurella multocida</i> .....	14

2.3.3.	Genómica comparativa en <i>Pasteurella multocida</i> .....	15
2.4.	Análisis pangenómico .....	16
2.4.1.	Antecedentes .....	16
2.4.2.	Definición del pangenoma .....	17
2.4.3.	Genoma core .....	17
2.4.4.	Genoma accesorio.....	18
2.4.5.	Pangenoma abierto y cerrado .....	19
2.5.	Diversidad genómica y adaptación.....	20
2.6.	Filogenómica .....	21
<b>3.</b>	<b>HIPÓTESIS Y OBJETIVOS.....</b>	<b>22</b>
3.1.	Hipótesis General.....	22
3.2.	Objetivo General .....	22
3.3.	Objetivos Específicos.....	22
<b>4.</b>	<b>MATERIALES Y MÉTODOS.....</b>	<b>23</b>
4.1.	Materiales.....	23
4.1.1.	Secuencias genómicas de <i>P. multocida</i> .....	23
4.1.2.	Base de datos .....	23
4.1.2.1.	Base de datos MLST RIRDC de <i>Pasteurella multocida</i> .....	23
4.1.2.2.	Base de datos de proteínas nr .....	23
4.1.2.3.	Base de datos de COG y CDD .....	23
4.1.3.	Softwares .....	24
4.1.3.1.	Sistema Operativo: BIOLINUX: GNU/LINUX - UBUNTU. ....	24
4.1.3.2.	Lenguajes de programación.....	24
4.1.3.2.1.	Bash.....	24
4.1.3.2.2.	Phyton.....	24
4.1.3.2.3.	R.....	25
4.1.3.3.	Programas .....	25
4.1.3.3.1.	MUSCLE .....	25
4.1.3.3.2.	JModelTest2 .....	25
4.1.3.3.3.	Prokka.....	25
4.1.3.3.4.	OrthoMCL .....	25
4.1.3.3.5.	MySQL.....	26
4.1.3.3.6.	BLAST.....	26
4.1.3.3.7.	R 3.4.0 y RStudio .....	26
4.1.3.3.8.	Phipack .....	27
4.1.3.3.9.	trimAL.....	27
4.1.3.3.10.	FASconCAT .....	28

4.1.3.3.11. ProtTest 3.3 .....	28
4.1.3.3.12. RAxML .....	28
4.1.3.3.13. RPS-BLAST .....	28
4.1.4. Hardware.....	28
4.1.4.1. Workstation.....	28
4.2. Métodos.....	29
4.2.1. Elección de secuencias genómicas de <i>Pasteurella multocida</i> .....	29
4.2.2. Re- anotación de los genomas de <i>Pasteurella multocida</i> .....	29
4.2.3. Predicción de familias génicas ortólogas: Recuperación de genes ortólogos.....	29
4.2.4. Perfil y tamaño del pangenoma de <i>Pasteurella multocida</i> .....	30
4.2.5. Anotación funcional: Análisis de agrupamiento de grupos ortólogos (COG) de proteínas.....	31
4.2.6. Análisis Filogenético .....	31
4.2.6.1. Filogenias basado en secuencia.....	31
4.2.6.1.1. Filogenia MLST.....	31
4.2.6.1.2. Análisis filogenómico del genoma core.....	32
4.2.6.1.3. Análisis de recombinación.....	32
4.2.6.2. Filogenia basada en contenido génico (árbol pangenómico).....	33
4.2.7. Análisis de la diversidad de genoma accesorio .....	33
4.2.8. Pipeline .....	33
<b>5. RESULTADOS.....</b>	<b>35</b>
5.1. Genomas seleccionados para el análisis pangenómico .....	35
5.2. La re- anotación de genomas muestra una diversidad en contenido genético. 36	
5.3. Agrupación de familias génicas y características generales del pan-genoma	39
5.4. Plot de pangenoma y genoma core.....	41
5.5. Anotación funcional del pangenoma de <i>Pasteurella multocida</i> .....	43
5.6. Análisis Filogenómico .....	45
5.6.1. Análisis filogenético a partir de MLST in <i>silico</i> .....	45
5.6.2. Análisis filogenético del Genoma core .....	46
5.6.3. Reconstrucción filogenética del Pangenoma .....	49
5.7. Diversidad del genoma accesorio.....	51
5.8. Identificación de genes relevantes del genoma accesorio que agrupan cepas relacionadas a enfermedad específica .....	53
5.9. Descripción funcional de genes presentes diferencialmente entre cepas asociadas a enfermedades específicas .....	55
5.9.1. Proteínas presentes de manera diferencial en cepas asociadas a septicemia hemorrágica .....	55

5.9.2. Proteínas presentes de manera diferencial en cepas asociadas a pasteurellosis neumónica.....	59
5.9.3. Proteínas presentes de manera diferencial en cepas asociadas a cólera aviar 62	
5.9.4. Proteínas presentes/ausentes de manera diferencial en cepas asociadas a pasteurellosis en humanos .....	67
<b>6. DISCUSIÓN .....</b>	<b>69</b>
6.1. El análisis pangenómico muestra una amplia variación en el contenido genético de <i>P. multocida</i> .....	69
6.2. El análisis filogenético revela que cepas asociadas a enfermedades específicas conforman clados separados. ....	70
6.3. El genoma accesorio presenta genes distribuidos diferencialmente en cepas de <i>P. multocida</i> asociadas a distintas enfermedades.....	73
6.4. La presencia de proteínas con función de transporte y metabolismo de carbohidratos pueden conferir capacidad para colonizar nuevos hospederos .....	75
<b>7. CONCLUSIONES .....</b>	<b>78</b>
<b>8. RECOMENDACIONES .....</b>	<b>79</b>
<b>9. REFERENCIA BIBLIOGRÁFICAS.....</b>	<b>80</b>
<b>10. GLOSARIO .....</b>	<b>90</b>



## LISTA DE TABLAS

**TABLA 1:** Lista de los 50 genomas de *P. multocida* usadas en el presente estudio, distribuidos por hospedero y enfermedad.

**TABLA 2:** Características de los genomas de las 50 cepas de *P. multocida* usadas en este estudio.

**TABLA 3:** Clústeres de cepas generado por kpax2 a partir de la matriz de presencia ausencia del genoma accesorio de *P. multocida*.

**TABLA 4:** Detalle de las proteínas presentes en cepas de *P. multocida* asociadas a septicemia hemorrágica.

**TABLA 5:** Detalle de las proteínas presentes en cepas de *P. multocida* asociadas a procesos neumónicos.

**TABLA 6:** Detalle de las proteínas presentes en cepas de *P. multocida* asociadas a cólera aviar.

## LISTA DE FIGURAS

**FIGURA 1:** Elaboración de un pangenoma.

**FIGURA 2:** Perfil del pangenoma. Un pangenoma puede ser abierto o cerrado.

**FIGURA 3:** Pipeline de la metodología usada en la presente investigación.

**FIGURA 4:** Características del pangenoma de *P. multocida*.

**FIGURA 5:** Perfil del *core* y pan-genoma de *P. multocida*.

**FIGURA 6:** Anotación funcional de los genomas *core*, accesorio y único de *P. multocida*.

**FIGURA 7:** Árbol filogenético MLST de *P. multocida*.

**FIGURA 8:** Árbol filogenómico de *P. multocida* del genoma *core*.

**FIGURA 9:** Árbol pangenoma de *P. multocida* y Tanglegrama.

**FIGURA 10:** Heatmap del pangenoma de *Pasteurella multocida*.

**FIGURA 11:** Heatmap de agrupamiento de cepas calculada por K-pax2 del genoma accesorio.

**FIGURA 12:** Heatmap que muestra la presencia de proteínas de manera diferencial en las cepas de septicemia hemorrágicas.

**FIGURA 13:** Heatmap que muestra la presencia de proteínas de manera diferencial en las cepas asociadas a pasteurelosis neumónica.

**FIGURA 14:** Heatmap que muestra la presencia de proteínas de manera diferencial en las cepas asociadas a cólera aviar.

**FIGURA 15:** Heatmap que muestra la presencia de proteínas de manera diferencial en las cepas asociadas a pasteurelosis en humanos.

## RESUMEN

*Pasteurella multocida* es una bacteria patógena responsable de un amplio rango de enfermedades en distintos hospederos animales incluyendo humanos. A pesar de su impacto en la salud y economía, todavía no se conocen las diferencias genéticas que pueden ser responsables de la variada patogénesis de *P. multocida*. En el presente trabajo, utilizando herramientas bioinformáticas, se realizó un análisis pangenómico y filogenómico para estudiar la diversidad genómica de esta bacteria y determinar si las cepas de distintos hospederos y enfermedades contienen diferencias en cuanto a contenido genético. Según el análisis, *Pasteurella multocida* posee un pangenoma abierto de 4881 genes y un *genoma core* de 1205 genes (25%). El *genoma* accesorio y único (75%) contienen mayormente secuencias profago y proteínas de función desconocida, junto con la alta presencia de recombinación en el *genoma core* (45%) son los principales generadores de diversidad mediante transferencia horizontal de genes y recombinación homóloga. El análisis filogenómico del *genoma core* y pangenoma muestra el agrupamiento de cepas asociadas a enfermedades específicas sugiriendo una especialización en esta bacteria con presencia de genes de manera diferencial en cepas asociadas a enfermedades como septicemia hemorrágica, cólera aviar, pasteurelosis neumónica y relacionados a infecciones en humanos. Las cepas de *P. multocida* asociadas a cólera aviar poseen operones relacionados a metabolismo de carbohidratos como L-fucosa, D-alosa, citrato, L-arabinosa, tagatosa y galactitol, en tanto que cepas asociadas a infecciones neumónicas presentan un operón de trehalosa, las cepas asociadas a septicemia hemorrágica poseen genes de biosíntesis de lipopolisacáridos, adherencia celular y defensa a macrófagos, y algunas cepas relacionadas a humanos carecen de conjuntos de genes asociados principalmente a transporte y metabolismo de carbohidratos. Estas diferencias en contenido y función genética pueden ser responsables de la variada patogénesis de *P. multocida* relacionado con la colonización e invasión en los primeros momentos de la infección.

Palabras clave: *Pasteurella multocida*, pangenoma, filogenómica, *genoma accesorio*, transferencia horizontal de genes.

## ABSTRACT

*Pasteurella multocida* is a Gram-negative pathogen that infects a wide range of animals including humans. Despite its impact on the health and economy, still remains unclear the genetic basis of its capacity to produce different clinical manifestations. In the present work, using bioinformatic tools, a pangenomic and phylogenomic analysis was carried out to study the genomic diversity of this bacterium and to determine if the strains of different hosts and diseases contain differences in their genetic content. According to the analysis, *P. multocida* has an open pangenome of 4881 genes and a core genome of 1205 genes (25%). The accessory and unique genome (75%) contains mostly profane sequences and proteins of unknown function, together with the high presence of recombination in the core genome (45%) are the main generators of diversity through gene transfer and homologous recombination. The phylogenomic analysis of the core genome and pangenome shows the grouping of strains associated with specific diseases suggesting a specialization in this bacterium with the presence of genes differentially in strains associated with diseases such as hemorrhagic septicemia, poultry cholera, pneumonic pasteurellosis and related to infections in humans. Strains of *P. multocida* associated with avian cholera have operons related to carbohydrate metabolism such as L-fucose, D-allose, citrate, L-arabinose, tagatose, and galactitol differentially; strains related to pneumonic pasteurellosis have remarkably a trehalose operon; whereas strains associated with hemorrhagic septicemia have genes responsible for lipopolysaccharide biosynthesis and defense against macrophages; and some strains related to humans lack genes associated mainly with carbohydrate metabolism and transport. These differences in gene content and function may be responsible for the specialization of *P. multocida* strains to produce a specific disease indicating a great importance of carbohydrate metabolism and others genes that may be involved in colonization of different environments within the host in early pathogenesis.

Key words: *Pasteurella multocida*, pangenome, phylogenomics, accessory genome, horizontal gene transfer

## 1. INTRODUCCIÓN

*Pasteurella multocida* es una bacteria Gram-negativa que está asociada a un amplio espectro de enfermedades en animales domésticos y silvestres incluido el humano produciendo cuantiosas pérdidas económicas y amenazando la salud pública (Wilson y Ho, 2013). En el Perú, la neumonía pasteurelósica en crías de alpaca es una de las principales causas de mortalidad que limita el desarrollo alpaquero en la región andina (Rosadio *et al.*, 2011). A pesar de los avances alcanzados en el entendimiento de la variada patogenicidad en diferentes hospederos de *P. multocida* por tecnologías de tipificación molecular y secuenciamiento genómico (Blackall *et al.*, 2000, Boyce *et al.*, 2012), todavía no resulta claro las diferencias entre distintas cepas de *P. multocida* y los casos clínicos que producen, así como los genes que pueden estar involucrados en esa especificidad, siendo necesario un análisis genómico más amplio que permita entender las bases moleculares de la variada patogenicidad de *P. multocida*. (Okay y Kizildogan, 2015, Du *et al.*, 2016).

El avance de la tecnología de secuenciamiento ha permitido el incremento en decenas de las secuencias genómicas de *P. multocida* que pueden ser analizadas usando nuevas herramientas bioinformáticas. Uno de esos métodos es el análisis pangenómico que explora las diferencias en el contenido genético de varios genomas (genes *core*, genes accesorios y genes únicos), las cuales son producto de mutaciones, recombinación y sobre todo de la transferencia genética horizontal y que pueden conducir a la adaptación a nichos específicos y la habilidad para colonizar nuevos hospederos (Tettelin *et al.*, 2008, McInerney *et al.*, 2017).

En el presente trabajo se analizó el pangenoma de *P. multocida* asociadas a distintos hospederos con el objetivo de estudiar la diversidad genómica y las diferencias en cuanto a contenido genético (presencia/ausencia) en cepas asociadas a distintos hospederos y enfermedades que favorecen un mayor entendimiento de la variada patogénesis de esta bacteria.

## **2. MARCO TEÓRICO**

### **2.1. Enfermedades asociadas a *P. multocida***

#### **2.1.1. Definición, impacto económico y amenaza zoonótica**

La bacteria *Pasteurella multocida* normalmente forma parte de la flora microbiana de la orofaringe de muchas especies de vertebrados. Sin embargo, también ha sido identificado como agente causal de múltiples enfermedades en un amplio rango de animales incluyendo humanos (Wilson y Ho *et al.*, 2013). *P. multocida* puede desenvolverse como agente primario de enfermedades que incluyen la septicemia hemorrágica (SH) en ungulados, cólera aviar (CA) en varias especies de aves, rinitis atrófica (RA) en cerdos, y coriza en conejos. También puede estar muy involucrada en la etiología de otras enfermedades como las infecciones de tracto respiratorio inferior como las neumonías en bovinos, cerdos, ovinos y camélidos, aunque requieren de otros factores para que la condición se desarrolle (Wilkie *et al.*, 2012). En humanos, las infecciones debido a *P. multocida*, han sido asociadas con la exposición a animales y usualmente involucran tejidos blandos después de mordeduras o rasguños de gatos o perros (Giordano *et al.*, 2015).

El término pasteurelosis es usado para describir las infecciones en mamíferos y aves causadas por *P. multocida*. A nivel mundial la pasteurelosis causa enormes pérdidas económicas en la industria ganadera, se calcula que las infecciones respiratorias en el ganado vacuno, principalmente pasteurelosis neumónica, produce pérdidas anuales que llegan a los mil millones de dólares en los Estados Unidos (McVey, 2009). Asimismo, la septicemia hemorrágica es la principal enfermedad de importancia económica en Asia y África en la producción de ganado vacuno y búfalos que impacta en la economía de los ganaderos pobres de la región (Shivachandra *et al.*, 2011). Por otro lado, el cólera aviar tiene un fuerte impacto económico en producciones aviares de grande y pequeña escala, esta enfermedad es de ocurrencia global y es capaz de infectar a todas las especies de aves, recientemente se ha demostrado que brotes recurrentes de cólera aviar puede conducir a la extinción de una colonia del ave silvestre, eider común (*Somateria mollissima: Anatidae*), en el ártico (Descamps *et al.*, 2012).

En el Perú la neumonía aguda es una de las principales causas de muerte en alpacas jóvenes donde *P. multocida* es principalmente aislada de los casos fatales, afectando la economía de los criadores andinos (Rosadio *et al.*, 2011; Guzmán *et al.*, 2013). En humanos, la pasteurelosis está asociada con una significativa morbilidad debido a complicaciones en heridas por mordeduras o rasguños de animales, llegando incluso de manera excepcional a producir la muerte. Aproximadamente el 1% (300,000) de las visitas anuales atendidas en emergencia en los EE.UU. son por heridas producto de mordeduras o rasguños de animales y *Pasteurella* es aislada del 50% y 75% de mordeduras de perros y gatos, respectivamente (Christenson *et al.*, 2015). Considerando la alta prevalencia de *P. multocida* como parte de la microbiota de animales domésticos, se puede considerar la transmisión zoonótica de *P. multocida* como un serio riesgo de salud pública (Wilson y Ho *et al.*, 2013).

A continuación, se describe resumidamente las manifestaciones clínicas y patogénesis de las enfermedades producidas por *P. multocida*.

### **2.1.2. Septicemia hemorrágica (SH)**

La septicemia hemorrágica es una enfermedad sistémica altamente mortal que afecta principalmente al ganado vacuno y búfalos, siendo prevalente en Asia y África. Las cepas de *P. multocida* involucradas principalmente son de tipo capsular B (Asia y Europa) y E (África) pero también se han aislado cepas capsulares tipo A (Wilkie *et al.*, 2012; Jabeen *et al.*, 2017).

La infección se inicia presumiblemente luego del ingreso del microorganismo al tejido tonsilar. Las bacterias se multiplican en el tejido blando peritonsilar. Los microorganismos son capaces de crecer rápidamente cuando se han establecido en tejidos blandos. La combinación de necrosis de tejido, inducción de citoquinas a nivel local y sistémico, el desarrollo de endotoxemia y bacteriemia (producidas por LPS de la bacteria) inducen los signos clínicos de una enfermedad aguda rápidamente fulminante (Shivachandra *et al.*, 2011).

Los órganos involucrados durante el proceso de la enfermedad son el tracto respiratorio, seguido de amígdalas, nódulos linfáticos del tracto respiratorio

superior y nasofaringe, también se ha reportado de manera específica lesiones en el sistema nervioso de búfalos (Marza *et al.*, 2017). Los signos clínicos son fiebre, letargia, y edemas con abundante salivación, lagrimeo, y secreción nasal, seguido al poco tiempo de dificultad respiratoria, shock séptico con hemorragia generalizada, llevando a la muerte dentro de 1 a 3 días. El tratamiento con antibiótico puede ser efectivo en estados tempranos de la enfermedad, pero desde que los signos clínicos agudos de sepsis se manifiestan, la mortalidad es casi de 100% (Wilkie *et al.*, 2012)

### **2.1.3. Cólera aviar (CA)**

El cólera aviar es una enfermedad contagiosa que afecta aves domésticas y silvestres. Todas las especies de aves parecen ser susceptibles, aunque algunas especies como el pavo son particularmente vulnerables a diferencia de los pollos que son más resistentes. La enfermedad es producida en mayor medida por cepas de tipo capsular A, aunque también se aíslan tipos capsulares D y F en pavos y se expresa en una amplia variedad de manifestaciones clínicas que van desde enfermedades sistémicas hiperagudas dominado por una alta mortalidad, a infecciones crónicas localizadas. El CA es considerada de significancia económica para la mayoría de tipos de aves de corral y el control depende principalmente de la vacunación y manejo de bioseguridad apropiada (Christensen y Bisgard, 2000).

La principal ruta de infección para *P. multocida* es el tracto respiratorio. Sin embargo, el aislamiento de la bacteria en casos de salpingitis y peritonitis indica que otras membranas mucosas pueden servir como puerta de entrada. Además, el microorganismo puede entrar al hospedero a través de heridas cutáneas. Después de la colonización del tracto respiratorio superior, *P. multocida* puede propagarse subsecuentemente a los pulmones, y/o múltiples tejidos, particularmente hígado y riñón, seguido de bacteriemia y septicemia (Wilkie *et al.*, 2012; Lee *et al.*, 2000).



#### **2.1.4. Rinitis atrófica (RA)**

Es un síndrome que puede incluir distorsión facial, destrucción del hueso de los cornetes y crecimiento retardado de cerdos jóvenes. Ha sido reconocido como una condición común de cerdos domésticos en todos los países. Los efectos clínicos son exclusivamente atribuidos a una dermonecrotoxina, referida como PMT (por “*Pasteurella multocida* toxin”), la cual está asociada con ciertas cepas de *P. multocida*, principalmente el tipo capsular D, y en menor medida el tipo capsular A (Wilkie *et al.*, 2012).

La PMT es una proteína de 146 kDa y actúa como un potente mitógeno para tipos celulares como osteoclastos y fibroblastos. En cerdos jóvenes, la administración de la PMT purificada sobre el cornete nasal es suficiente para causar rinitis atrófica. La toxina produce la enfermedad interfiriendo en la biogénesis del hueso promoviendo la diferenciación y proliferación de osteoclastos resultando en una reabsorción del hueso incrementada, mientras inhibe la diferenciación de osteoblasto y la regeneración del hueso por otro lado (Chakraborty *et al.*, 2017)

El principal signo clínico de la rinitis atrófica incluye distorsión maxilar, desviación del septo nasal, erosión y ausencia de cornetes nasales, y posiblemente, disminución en la tasa de crecimiento, aunque el último efecto puede estar más asociado a las prácticas de manejo que a la enfermedad. Las lesiones del tejido óseo son permanentes, aunque la mucosa puede repararse. La mayoría de los cerdos son sacrificados a una edad relativamente joven (Horiguchi, 2012).

#### **2.1.5. Coriza**

*P. multocida* es responsable de una de las enfermedades más serias en conejos que causa considerables pérdidas económicas en la producción. Puede causar la muerte en conejos o una infección local como rinitis, otitis media, conjuntivitis, y neumonía. Estos casos están asociados con la colonización del tracto respiratorio superior por *P. multocida* principalmente de tipo capsular A, aunque el tipo D también ha sido aislado (Massacci *et al.*, 2018).

El primer signo clínico de rinitis es una secreción nasal serosa. En algunos conejos, las bacterias se propagan a los pulmones donde pueden causar una pulmonía localizada crónica o una neumonía fibrinopurulenta severa, los signos clínicos incluirán tos, fiebre, dificultad respiratoria leve o severa y eventualmente la muerte (Wilkie *et al.*, 2012).

#### **2.1.6. Pasteurelosis neumónica**

Las infecciones de tracto respiratorio inferior (ITRI) o pasteurelosis neumónica, asociados con *P. multocida* pueden manifestarse como condiciones agudas fulminantes (por ejemplo, la fiebre de embarque o “shipping fever” del ganado vacuno) o crónicas, como la llamada “neumonía enzootica” en cerdos, terneros y ovejas. *P. multocida* es la especie más comúnmente aislada de neumonías de cerdos y alpacas (Wilkie *et al.*, 2012; Guzman *et al.*, 2013). Pero no es la única, *Mannhemia haemolytica*, *Mycoplasma sp.* e *Histophilus somni* también contribuyen en la enfermedad. Las cepas de *P. multocida* aisladas de pulmones son de tipo capsular A, en menor medida de tipo D y ocasionalmente de tipo F (Dabo *et al.*, 2008).

En la mayoría de casos, *P. multocida* actúa como un patógeno oportunista. Las condiciones que predisponen a la infección son pobremente definidas, pero el estrés fisiológico, sumado a infecciones virales y otras bacterias ha sido reconocido como contribuyente del establecimiento de *P. multocida* en el tracto respiratorio inferior de bovinos. De cómo y porqué *P. multocida* pasa de ser una bacteria comensal de tracto respiratorio superior a patogénica e invadir el pulmón, es todavía desconocido.

#### **2.1.7. Pasteurelosis como infección zoonótica**

La infección en humanos por *Pasteurella* es producida por mordeduras de animales, principalmente de perros y gatos. En las heridas, *P. multocida* es el patógeno predominantemente encontrado, especialmente en casos severos. La mayoría de aislados son de los serogrupos A y D (Giordano *et al.*, 2015). Las heridas producidas por mordedura desarrollan hinchazón (edema), celulitis (difuso, con inflamación localizada y dolor), y exudado purulento/supurativo con

sangrado en el sitio de la herida. La cantidad de leucocitos y neutrófilos son típicamente altas en el sitio de la infección, y la inflamación se desarrolla rápidamente. En los casos más severos, la pasteurelosis puede progresar rápidamente a bacteriemia y otras complicaciones como osteomielitis, endocarditis y meningitis. Estas complicaciones son más comunes en neonatos, ancianos e individuos inmunocomprometidos en quienes puede ser fatal (Wilkie *et al.*, 2012).

## **2.2. *Pasteurella multocida***

### **2.2.1. Historia y posición taxonómica**

En 1881 Luis Pasteur identifica el agente causal del cólera aviar en pollos. Desde entonces *Pasteurella multocida*, bautizada así en honor a su descubridor, ha sido identificada como el agente etiológico de muchas otras enfermedades. Debido a esa capacidad, su nombre de especie, *multocida*, puede traducirse literalmente como “multi-asesina” (Harper *et al.*; 2006).

Su nombre fue propuesto en 1938 para incluir a “todos los organismos no hemolíticos, productores de indol” (Rosenbusch y Merchant, 1938), pero su posición taxonómica fue definida formalmente por Mutters *et al.* (1985) con el cual quedó definido que *Pasteurella multocida* incluye tres subespecies: *P. multocida* subsp. *multocida*, *P. multocida* subsp. *septica*, y *P. multocida* subsp. *gallicida* y junto con otras especies como *P. avium* y *P. canis* pertenece al género *Pasteurella* y a la familia *Pasteurelacea* (Christensen y Bisgaard, 2006).

### **2.2.2. Identificación de *P. multocida***

#### **2.2.2.1. Identificación fenotípica de *P. multocida*: caracterización microbiológica y bioquímica**

*P. multocida* es Gram-negativa, cocobacilo, no motil, puede crecer bajo condiciones aeróbicas, anaeróbicas o en microaerofilia y lo hacen óptimamente entre 37°C y 41°C. Generalmente son cultivadas en agar sangre donde forman colonias redondas (1 – 2mm), de color gris claro, no hemolíticos (Dziva *et al.*, 2008). Un amplio rango de pruebas bioquímicas se realiza para la identificación

de *P. multocida*: principalmente por reacción positiva de ornitina carboxilasa, indol, formación ácida de (-)-D-manitol, y falta de producción de maltosa y dextrina (Christensen y Bisgaard, 2006). Sin embargo, la identificación fenotípica no es absoluta, una alta variabilidad fenotípica se ha reportado para *P. multocida* que hace difícil su clasificación por subespecies. La solución adecuada para la identificación definitiva es la combinación de pruebas fenotípicas y genéticas (Boyce *et al.*, 2010).

#### **2.2.2.2. Identificación genotípica de *Pasteurella multocida***

Se han desarrollado pruebas de PCR (Reacción en Cadena de la Polimerasa) para distinguir *P. multocida* de otras especies bacterianas. La más usada está dirigida a amplificar el gen *kmt1*, que codifica una esterasa/lipasa. Esta es capaz de amplificar en todas las cepas de *P. multocida* incluyendo cada una de sus subespecies pero no en otras especies del mismo género (Townsend *et al.*, 1998; Boyce *et al.*, 2010).

La secuencia del gen de ARNr 16S es ampliamente usado para la identificación de especies bacterianas en base a la comparación de secuencia y construcción de árboles filogenéticos (Kim *et al.*, 2014). En *P. multocida* la comparación de secuencia del fragmento de 1 468 pb del gen ARNr 16S tiene una homología de 99.8-99.9% dentro de la especie (Dey *et al.*, 2007). En tanto, se ha descrito una identidad de 97.4 % con *Pasteurella canis* y 93.7 % con *Pasteurella avium* (Davies, 2004).

#### **2.2.3. Tipificación de *Pasteurella multocida***

##### **2.2.3.1. Tipificación serológica**

Serológicamente, *P. multocida* es diferenciada según serogrupo y serotipo. La clasificación serológica basada en antígenos capsulares reconoce cinco serogrupos: A, B, D, E y F. La constitución capsular para serogrupo A (ácido hialurónico), D (heparina) y F (condroitina) (Harper *et al.*, 2012). El segundo método serológico, desarrollado por Heddlestone *et al.* (1972), es un test de precipitación en gel de difusión que reconoce 16 serotipos de LPS, aunque estas

no cubran la completa diversidad de LPS observada en cepas de campo (Harper y Boyce, 2017). El sistema de serotipificación estándar para *P. multocida* es representada por ambas pruebas serológicas: con el tipo de cápsula seguido del tipo de LPS (por ejemplo, A:3, B:2).

#### **2.2.3.2. Tipificación molecular**

Los métodos de tipificación molecular son requeridos para investigaciones epidemiológicas dado que la serotipificación no tiene suficiente poder discriminatorio para caracterizar clones de *P. multocida* derivados de distintos brotes infecciosos (Christensen y Bisgaard, 2006).

Los métodos basados en PCR como RAPD (*Random Amplification Polimorphic DNA*), ERIC-PCR (*Enterobacterial Repetitive Intergenic Consensus-PCR*) o REP-PCR (*Repetitive Extragenic Polymorphism-PCR*) han sido usados para separación genotípica de aislados de varios hospederos y brotes (Blackall y Mifflin, 2000; Dziva *et al.*, 2001; Rimac *et al.*, 2017). Asimismo, la PFGE (Electroforesis en Gel de Campo Pulsado) ha demostrado ser discriminatoria de aislados estrechamente relacionados y ha sido útil en el estudio epidemiológico de diversos brotes (Gunawardana *et al.*, 2000; Dziva *et al.*, 2008).

Por otro lado, el MLST (*Multilocus Sequencing Typing*), es método que caracteriza aislados por medio de secuenciamiento de fragmentos de siete genes *housekeeping*, tiene ventaja sobre los métodos RAPD y PFGE en términos de reproducibilidad y portabilidad en el análisis de cepas, aunque el uso de genes *housekeeping* puede limitar el poder discriminatorio y reducir su utilidad para investigaciones a nivel epidemiológico (Subaaharan *et al.*, 2010). Hotchkiss *et al.* (2011) consideran que MLST es la técnica más apropiada para estudios globales y de largo término. Esta viene siendo usada ampliamente para estudiar poblaciones de *P. multocida* aislados de diferentes especies animales, actualmente existen dos bases de datos de MLST: RIRDC y Multi hospedero (disponible en <https://pubmlst.org/pmultocida>).

#### **2.2.4. Factores de virulencia**

Muchos factores de virulencia de *P. multocida* han sido identificados, pero la contribución específica a la patogénesis todavía no es muy entendida en muchos casos. El único factor de virulencia crítico conocido es la PMT (toxina de *P. multocida*) que causa rinitis atrófica, y su acción está bien definida. Otras exotoxinas no han sido caracterizadas para otras enfermedades. Sin embargo, resulta claro que algunos productos de *P. multocida* que se describen a continuación son críticos para la virulencia (Boyce *et al.*, 2010)

##### **2.2.4.1. Cápsula**

*P. multocida* expresa una cápsula hidrofílica que inhibe la fagocitosis y la opsonización mediada por complemento. La ausencia de biosíntesis de la cápsula resulta en una cepa altamente atenuada que no resiste a los componentes del suero. En *P. multocida* el componente principal del tipo capsular A es el ácido hialurónico (Harper *et al.*, 2012).

##### **2.2.4.2. LPS (lipopolisacáridos)**

El lipopolisacárido (LPS) es un importante factor de virulencia que confiere resistencia al sistema complemento del suero. Mutantes con LPS truncado muestran una virulencia reducida y son susceptibles a ser eliminados por péptidos antimicrobianos del hospedero. En general el LPS es una endotoxina de la membrana externa de bacterias Gram negativas y actúa como un agente inmunoestimulante en animales (Harper y Boyce, 2017).

##### **2.2.4.3. Proteínas de membrana externa (PMEs), fimbrias y adhesinas**

La colonización del tejido del hospedero es un proceso esencial para el establecimiento de la infección. *P. multocida* posee una batería de genes fimbriales que incluyen *ptfA*, *fntA*, *fimA*, *flp1*, *flp2*, *hsf1* y *hsf2* que codifican proteínas similares a fimbria y fimbrillas de otras bacterias. Asimismo, las proteínas FLp, TadD y las proteínas hemaglutinina filamentosa PfhB1 y PfhB2 han sido identificadas por tener un rol en la virulencia. Por otro lado, las proteínas de membrana externa (OMP en inglés) también son importantes factores de

virulencia. Por ejemplo, OmpA actúan como adhesinas al unirse a proteínas de membrana externa de las células del hospedero. OmpH, una porina, y la lipoproteína E (PlpE) son antígenos protectivos asociados al tipo capsular A en casos de cólera aviar y pasteurelisis neumónica de bovinos. Oma37, Cp39 y Tbp también son otras proteínas de membrana involucradas en virulencia (Harper *et al.*, 2006, Hatfaludi *et al.*; 2010).

#### **2.2.4.4. Proteínas de metabolismo de hierro**

El hierro es esencial para la supervivencia de las bacterias. *P. multocida* posee mecanismos para el secuestro de hierro de hidróxido férrico precipitado vía sideróforos o de proteínas de hospedero ricos en hierro como transferrinas y grupos hemo. Ello requiere de ciertas proteínas de membrana externa para la captura de hierro. Un sistema conocido del sistema TonB consiste de proteínas TonB, ExbB, y ExbD juegan un rol importante en la adquisición de hierro. TbpA es una proteína de membrana de unión a transferrina y se encuentra en la mayoría de aislados de *P. multocida* en rumiantes (Harper *et al.*, 2006; Wilson y Ho, 2013).

#### **2.2.4.5. Sialometabolismo**

Una bacteria invasora puede utilizar el ácido siálico del hospedero como fuente de carbono y nitrógeno. *P. multocida* posee un sistema sialometabólico que comprende vías de captura, activación, transferencia y disimilación de ácido siálico. El ácido siálico es obtenido del hospedero usando sialidasas NanH, NanB que poseen distintas especificidades, *P. multocida* codifica también para dos sialiltransferasas capaz de transferir ácido siálico a componentes de superficie celular, el más conocido de estas enzimas es Pm0188, la otra es Pm0508 (Boyce *et al.*, 2010).

#### **2.2.4.6. Toxinas**

En general, las cepas de cólera aviar, septicemia hemorrágica y neumonía no poseen toxinas conocidas. La toxina dermonecrótica o toxina de *P. multocida* (PMT), expresada principalmente por el serogrupo D es responsable de los

signos clínicos y patológicos de la rinitis atrófica en cerdos. La toxina PMT se encuentra en un bacteriófago perteneciente a la familia *Siphoviridae*. PMT no tiene secuencia señal y ningún mecanismo de exportación, por lo que se sugiere que la fase lítica del bacteriófago puede mediar la liberación de la toxina (Boyce *et al.*; 2010; Horiguchi, 2012). La presencia del gen también ha sido encontrada a nivel bioinformático en el análisis genómico de la cepa UNMSM (serotipo A) aislado de alpacas (Hurtado *et al.*, 2017).

#### **2.2.5. Patogénesis molecular y celular**

Dado que *P. multocida* es responsable de un amplio rango de enfermedades de distintas presentaciones en numerosos hospederos es esperable que haya significantes diferencias en cuanto a la patogénesis de cada enfermedad. Existe cierta asociación de serotipos de *P. multocida* con determinadas enfermedades y hospedero. El cólera aviar es generalmente causado por tipos capsulares A y F, rinitis atrófica por cepas D, y septicemia hemorrágica por cepas tipo B (Harper *et al.*, 2012). Sin embargo, esta asociación no es completa, por ejemplo, el serotipo A:3 puede causar cólera aviar en pavos y pasteurelosis neumónica en ganado vacuno o alpacas (Okay y Kizildogan, 2015; Hurtado, 2017).

Se conoce de manera general la ruta de la infección de todas las enfermedades en la que *P. multocida* está involucrada: para la septicemia hemorrágica, la ruta de infección es el tejido tonsilar, en tanto para cólera aviar y pasteurelosis neumónica en ungulados, el tracto respiratorio parece ser el portal principal. Una vez que *P. multocida* ingresa al tejido o sangre, este puede multiplicarse inmediatamente (septicemia hemorrágica, neumonía o infecciones en heridas en humanos) o localizarse temporalmente en órganos tales como el hígado y riñón (aves) desde donde pasa después a la sangre (Boyce *et al.*, 2010). A pesar de estos avances, las bases moleculares y celulares de la asociación de cepas por hospedero o enfermedad permanecen desconocidos (Wilkie *et al.*, 2012). A continuación, se describen los procesos de patogenicidad conocidos hasta el momento para *P. multocida*.



El sistema inmune innato es una de las primeras barreras que una bacteria debe sortear. Este sistema comprende muchos mecanismos de defensa, como fagocitosis, sistema de complemento, actividad de péptidos antimicrobianos y secuestro de hierro libre. La cápsula y el lipopolisacárido (LPS) son los principales componentes de membrana y son importantes para que el patógeno sea fagocitado o matado por el sistema de complemento (Harper *et al.*, 2012). Recientemente se ha descrito que los LPS de *P. multocida* B:2 pueden estar involucrados en la despolimerización de actina lo que facilitaría la invasión en células endoteliales de aorta bovina (CEAB) (Yap *et al.*, 2017).

*P. multocida* posee muchas proteínas involucradas en la obtención de hierro que se encuentran en baja concentraciones en células de aves y mamíferos. El genoma de la cepa de referencia, Pm70, posee más de 50 proteínas con funciones predichas de adquisición y transporte de hierro (May *et al.*, 2001). La redundancia de función y proteínas parálogas relacionados con la captura de hierro es una característica resaltante de la biología de *P. multocida*, por ejemplo, los genes *hgbA* y *hgbB* son proteínas parálogas de unión a hemoglobina. Algunas proteínas de adquisición de hierro han sido evaluadas como potenciales vacunas, pero sin ningún éxito, es probable que esto se deba al alto grado de redundancia de función (Boyce *et al.*, 2010).

Para el establecimiento de cualquier infección, la adhesión al tejido del hospedero es un requisito fundamental. Se cree que el constituyente capsular, ácido hialurónico puede estar involucrado en la adhesión a hospedero, esto se ha demostrado en macrófagos y monocitos de pavo. De la misma manera, la proteína de membrana OmpA, se une a fibronectina y heparina de la superficie de células de riñón de bovino (Dabo *et al.*; 2008), favoreciendo la adhesión celular. Se ha demostrado que variantes de esta proteína pueden estar relacionado con el grado de virulencia de la bacteria (Katoch *et al.*, 2014). La presencia de los genes *flp1*, *flp2* y locus *tadD* involucrados en la producción de fimbria es ubicua en *P. multocida* y están presente en los tipos capsulares A, B y D.

No se considera a *P. multocida* como un patógeno intracelular. Sin embargo, se ha demostrado que posee la habilidad de invadir a una variedad de tipos celulares. Las cepas virulentas A:3 y A:4 pueden ser capaces de invadir células de riñón de pavo. Lo mismo que cepas B:2 poseen la capacidad de invadir células endoteliales (Wilkie *et al.*, 2012). Recientemente se ha demostrado que esta capacidad es diferente entre los tipos A:3 y B:2, este último posee la capacidad de invadir y multiplicarse en células embrionarias de pulmón de bovino, mientras que las primera no. Esta capacidad puede estar relacionado con la capacidad de translocarse de la mucosa pulmonar a través de células endoteliales hasta la sangre para generar bacteriemia como última fase de infecciones septicémicas (Wilkie *et al.*, 2012).

### **2.3. Genómica bacteriana y de *Pasteurella multocida***

#### **2.3.1. Secuenciamiento genómico en bacterias**

El secuenciamiento de genomas bacterianos viene desarrollándose desde 1996. Durante este tiempo, el avance y abaratamiento de las tecnologías de secuenciación sumado a nuevos enfoques bioinformáticos de análisis de datos han transformado nuestro entendimiento de cómo las bacterias funcionan, evolucionan e interactúan con su hospedero y ambiente (Loman y Pallen, 2015).

#### **2.3.2. Genoma de *Pasteurella multocida***

En el 2001 se secuenció el primer genoma completo de *Pasteurella multocida* cepa Pm70 por la tecnología Sanger (May *et al.*, 2001). Desde entonces, con la aparición y masificación de las tecnologías de secuenciamiento de nueva generación se han descrito nuevas cepas de *P. multocida* (Liu *et al.*, 2012; Abrahante *et al.*, 2013; Peng *et al.*, 2016). Entre ellas el reciente secuenciamiento y caracterización del genoma de *P. multocida* cepa UNMSM aislada de un caso de neumonía de alpaca (Hurtado *et al.*, 2017).

Todos los genomas de *Pasteurella multocida* disponibles en el Genbank varían entre 2 - 2.4 Mbp de longitud y comprenden un genoma circular único con un contenido de G+C de 40 - 41% con más de 2000 secuencias de ADN codificante

(CDS). Muchas de estas secuencias genómicas se han usado en estudios para el descubrimiento de nuevos factores de virulencia (May *et al.*, 2001), la predicción de epítopes o antígenos inmunogénicos para su uso como vacunas mediante vacunología reversa (Maximiliano *et al.*, 2018) y estudios de genómica comparativa (Moustafa *et al.*, 2015).

### **2.3.3. Genómica comparativa en *Pasteurella multocida***

La genómica comparativa es la comparación del contenido genético de un organismo contra otro. Este enfoque puede ayudar a determinar la función de genes y regiones de secuencias codificantes de genomas, así como caracterizar y determinar relaciones filogenéticas y eventos evolutivos (Sivashankari y Shanmughavel, 2007). El análisis comparativo tiene como primer objetivo identificar las semejanzas y diferencias entre organismos (Binnewies *et al.*, 2006).

En los últimos 5 años, gracias al incremento de secuencias genómicas de *P. multocida*, varios trabajos de genómica comparativa se han llevado a cabo, logrando encontrar grupos de genes asociados a virulencia, islas de patogenicidad y profagos portadores de genes de resistencia, así como grupos de genes únicos en cepas asociados a septicemia hemorrágica o cólera aviar (Johnson *et al.*, 2013; Moustafa *et al.*, 2015; Yu *et al.*, 2016).

Entre los trabajos de genómica comparativa realizados hasta el momento encontramos los siguientes: Boyce *et al.* (2012) reportaron las diferencias entre los genomas de la cepa Pm70 (cepa no patógena aislada de pavo) y cepa 36950 (cepa patógena aislada de bovinos) en términos de sus características básicas. En tanto Johnson *et al.* (2013) compararon los genomas de las cepas aviares Pm70, X73 y P1059 reportando genes únicos en las cepas virulentas. En un interesante trabajo en 2015, Okay y Kurt analizaron 5 genomas de cepas de distintos hospederos evaluando el porcentaje de proteínas que mutan en una comparación de genomas de par en par reportando mutaciones en genes relacionados a patogenicidad especialmente en proteínas de membrana externa y distintos procesos metabólicos. Du *et al.* (2016) compararon 2 genomas de

cepas virulentas aisladas de bovinos con otras cepas conocidas (3690, HN06, 3480, Pm70) encontrando diferentes genes de virulencia que puede indicar variación en la patogenicidad y sugieren un análisis más profundo.

Los estudios genómicos realizados hasta el momento han logrado descubrir nuevos factores de virulencia asociados a las diferentes enfermedades que produce. Además, han demostrado que *P. multocida* posee una diversidad genómica intraespecífica importante que permite especular que genes específicos de diferentes cepas puedan jugar un rol importante en los mecanismos de patogenicidad variada de *P. multocida*. Sin embargo, Boyce *et al.*, (2012) basado en el análisis de 9 genomas de origen distinto (aviar, bovino y cerdo) encontraron que “no había una clara correlación entre los grupos filogenéticos y la predilección a hospedero y enfermedad”. Este resultado puede ser parcial debido a la cantidad de genomas analizados, una aproximación más extensiva, es decir con una mayor cantidad de genomas y nuevos enfoques pueden permitir dilucidar si *P. multocida* tiene una segregación por hospedero y enfermedad.

Recientemente, Hurtado, *et al.* (2018) utilizan 23 genomas completos de *Pasteurella multocida* identificando mediante un análisis filogenómico una diversificación para cepas asociadas a septicemia hemorrágica y pasteurelosis neumónica sugiriendo la existencia de genes involucrados en especialización, siguiendo esta línea todavía es necesario un análisis con mayor número de genomas y la identificación de aquellos genes involucrados en dicha especialización.

## **2.4. Análisis pangenómico**

### **2.4.1. Antecedentes**

El secuenciamiento masivo de genomas y la genómica comparativa ha cambiado nuestro entendimiento de lo que son las bacterias. Las bacterias se replican clonalmente entre generaciones, pero también son capaces de recombinar entre individuos y transferir horizontalmente su ADN mediante transformación, transducción y conjugación (Soucy *et al.*, 2015). Esto conduce a la ruptura del

desequilibrio de ligamiento que se produciría en las poblaciones clonales, debido a ello, la diversidad genómica de una bacteria es a menudo caracterizada por una larga variación en su repertorio génico (Bobay *et al.*, 2015; Rodríguez-Varela *et al.*, 2016).

La extensión de la genómica comparativa a un número mayor de secuencias genómicas abre la posibilidad de realizar profundos análisis sobre la diversidad genómica de una bacteria que permita inferir su repertorio genético total, genoma mínimo, estilo de vida, evolución y adaptación a nicho específico. Esto representa un desafío para el análisis comparativo que ya no debe encargarse de pocos sino de decenas, cientos y hasta miles de genomas de manera eficiente y precisa (Vernikos *et al.*, 2015).

#### **2.4.2. Definición del pangenoma**

En este contexto, el enfoque denominado *pan-genoma* se ha desarrollado para estudiar comparativamente múltiples genomas estableciendo nuevos conceptos y herramientas de análisis relacionados con la naturaleza y diversidad genómica de procariotas. El término pan-genoma fue acuñado hace poco más de una década por Tettellin *et al.* (2005) como el contenido genético total de una especie bacteriana. El pan-genoma puede dividirse en: el genoma core, que representa el conjunto de genes compartidos por todas las cepas de una especie y el genoma accesorio compuesta de genes presentes en algunas cepas, pero no en todas. En general, los genes del genoma core son responsables de los procesos celulares básicos y las principales características de la especie; mientras el genoma accesorio, influenciado por la transferencia horizontal, la ganancia y pérdida de material genético, contribuyen a la diversidad dentro de la especie y puede tener efecto adaptativo (Tettellin *et al.*; 2008).

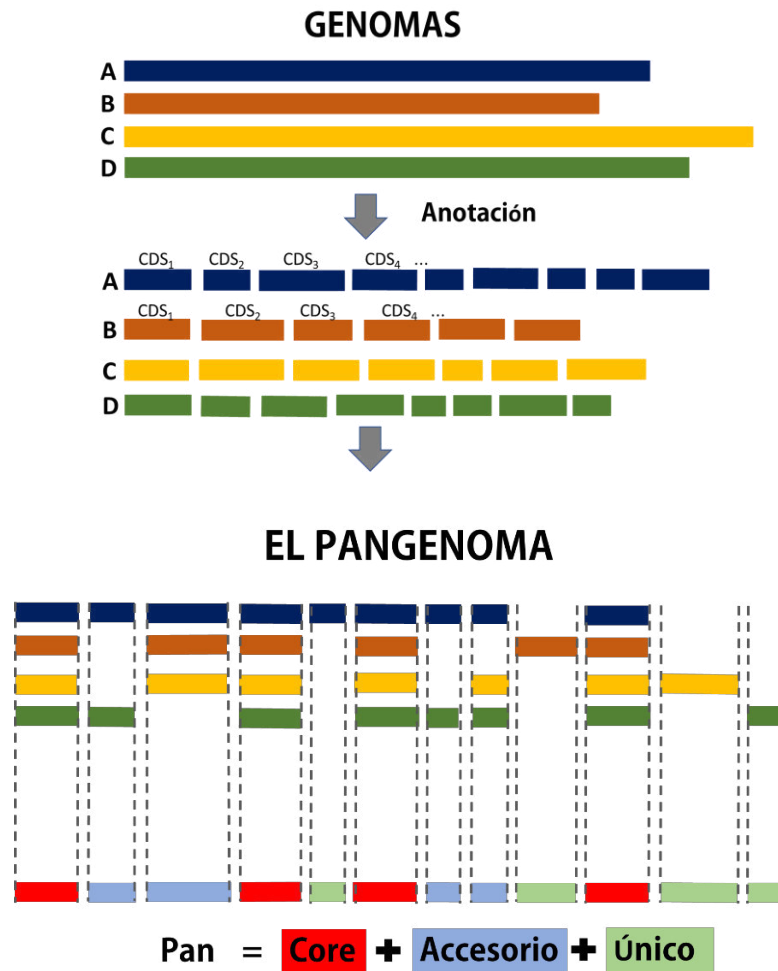
#### **2.4.3. Genoma core**

El genoma core contiene normalmente los genes que están asociados con el mantenimiento de los aspectos básicos de la biología de un organismo principalmente relacionados a metabolismo energético, replicación, traducción y mantenimiento de la homeostasis celular. Debido a ello el genoma core sufre

significante presión selectiva en relación a su función, el cual inhibe la ocurrencia de cambios drásticos (Tettellin *et al.*, 2008; Vernikos *et al.*, 2015).

#### **2.4.4. Genoma accesorio**

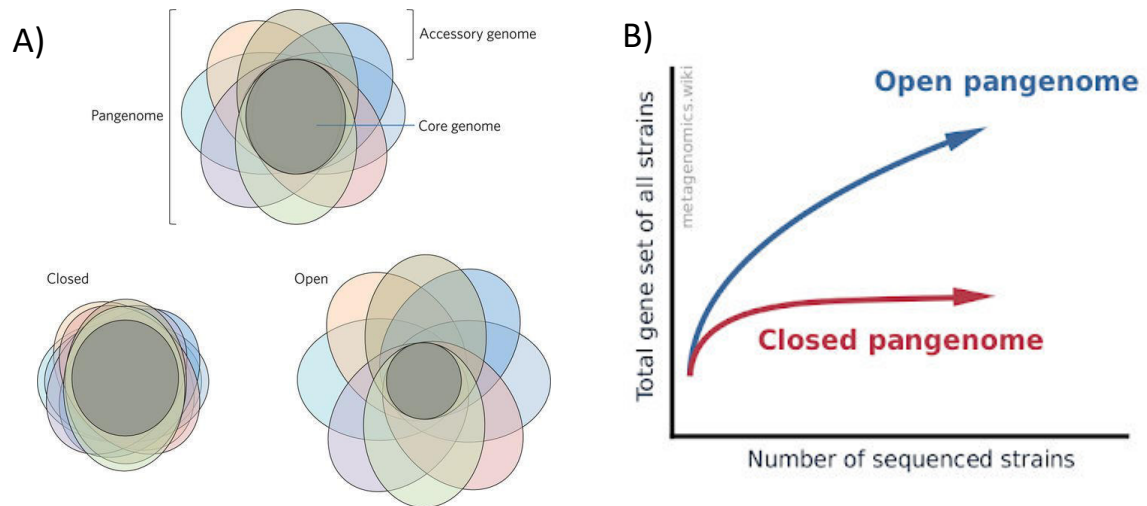
Todas las cepas de una especie bacteriana poseen un conjunto conservado de genes (core) y adicionalmente un conjunto de genes accesorios que es compartido por varias cepas, pero no por todas. La dinámica de ganancia de genes facilitada por diversos mecanismos de transferencia genética horizontal permite que diferentes cepas puedan poseer diferentes conjuntos de genes. Es probable que los genes accesorios constituyan un “reservorio” que puede ser transferido lateralmente para crear nuevas combinaciones. Se ha descrito que buena parte del genoma accesorio frecuentemente presenta genes con anotación no funcional o el conocimiento de estos genes es pobre o nula (Segerman, 2012). El genoma accesorio probablemente proporcione funciones que no son esenciales para las actividades básicas, pero pueden conferir ventajas selectivas que incluyen la adaptación a ambientes específicos (McInerney *et al.*, 2017).



**Figura 1: Elaboración del pangenoma.** Anotación de genomas, identificación de secuencias ortólogas y agrupamiento de familias ortólogas de genes. Pangenoma = core + accesorio + único. **(Diseño de elaboración propia)**

#### 2.4.5. Pangenoma abierto y cerrado

Algunas especies bacterianas tienen un pangenoma “abierto” que se expande cuando nuevos genomas son adicionados al pangenoma, mientras otras tienen pangenomas “cerrados” con diferencias genéticas muy pequeñas y que no cambia cuando nuevos genomas son adicionados (McInerney *et al.*, 2017) (Figura 2). La definición de un pangenoma como abierto o cerrado responde a la capacidad de la especie de adquirir nuevo material genético. La naturaleza abierta o cerrada de un pangenoma está relacionada al estilo de vida de la especie bacteriana (Rouli *et al.*, 2015).



**Figura 2: Perfil del Pangenoma. Un pangenoma puede ser abierto o cerrado.** A. El pangenoma abierto posee un contenido de genes accesorio más abundante que en un pangenoma “cerrado”. B. El pangenoma “abierto” se expande indefinidamente según el número de secuencias analizadas, en tanto uno “cerrado” alcanza un estado casi asintótico en donde el incremento de secuencias no aumenta el número total de genes (Imágenes tomadas de McInerney *et al.*, 2017)

## 2.5. Diversidad genómica y adaptación

Las diferencias en el contenido genético debido a la dinámica de ganancia y pérdida de material genético hacen que las cepas bacterianas relacionadas a un nicho específico posean grupos de genes exclusivamente presentes o ausentes en ellos; es decir, estos genes pueden ser la respuesta parcial de la adaptación local de diversos patógenos bacterianos a particulares hospederos o a la especialización patogénica (Wiedenbeck y Cohan, 2011; Gyles *et al.*, 2014). Se ha descrito que buena parte de los genes que se transmiten horizontalmente son adaptativos, es decir pueden conferir significantes beneficios en términos de mejor aprovechamiento de sustrato o recurso, nuevas capacidades metabólicas, tolerancia al *stress* o resistencia antimicrobiana (Vos *et al.*, 2015; McInerney *et al.*, 2017).



## 2.6. Filogenómica

Los análisis evolutivos permiten poner en perspectiva los estudios de genómica comparativa. A partir de ello podemos empezar a entender cómo e incluso porqué las similitudes o diferencias entre genomas, por ejemplo, la presencia o ausencia de genes, los patrones de sustitución en el ADN, y patrones globales de sintenia (orden conservado) entre especies. Estos análisis, a su vez, pueden ser usados para entender el metabolismo, la fisiología o patogenicidad. El análisis evolutivo a nivel genómico lo hace posible la filogenómica (Eisen y Fraser, 2003). La filogenómica es definida como la reconstrucción de la historia evolutiva de organismos tomando en cuenta todo el genoma. En este contexto, un “genoma” puede ser entendido como su secuencia de ADN, o su contenido genético (generalmente solo genes codificantes de proteínas) (Patané *et al.*, 2018). El genoma core suele utilizarse como *input* para la reconstrucción filogenética.

### **3. HIPÓTESIS Y OBJETIVOS**

#### **3.1. Hipótesis General**

Existen genes que están presentes diferencialmente en los genomas de las cepas de *Pasteurella multocida* relacionadas a hospedero y enfermedad específicas.

#### **3.2. Objetivo General**

Realizar un análisis pangenómico de la bacteria *Pasteurella multocida* e investigar si existen genes que están presentes diferencialmente en los genomas derivados de hospedero y enfermedad específica

#### **3.3. Objetivos Específicos**

- Caracterizar el pangenoma de *Pasteurella multocida*: genoma core, genoma accesorio y genes únicos.
- Realizar un análisis filogenómico de *Pasteurella multocida*.
- Realizar un análisis de recombinación del genoma core de *P. multocida*.

## **4. MATERIALES Y MÉTODOS**

### **4.1. Materiales**

#### **4.1.1. Secuencias genómicas de *P. multocida***

Las secuencias genómicas que se usaron en la presente investigación fueron descargadas de la base de datos del Genbank (Clark *et al.*, 2016) para genomas (<ftp://ftp.ncbi.nlm.nih.gov/genomes/>) de la especie *Pasteurella multocida*.

#### **4.1.2. Base de datos**

##### **4.1.2.1. Base de datos MLST RIRDC de *Pasteurella multocida***

La base de datos MLST para *Pasteurella multocida* es una plataforma administrada por BIGDdb Genomics, esta contiene las secuencias de alelos de los perfiles de MLST, así como información de procedencia y epidemiología. El esquema RIRDC MLST fue desarrollado por el Animal Research Institute de Australia con financiamiento de Rural Industries Research and Development Corporation (RIRDC). Se encuentra disponible en <https://pubmlst.org/pmultocida/>. Recientes trabajos de caracterización genética y genómica de *P. multocida* han utilizado este esquema (Hotchkiss *et al.*, 2011; Moustafa *et al.*, 2015).

##### **4.1.2.2. Base de datos de proteínas nr**

La base de datos nr del Genbank posee el registro de todas las secuencias de proteínas sometidas por laboratorios individuales y proyectos de gran escala. Es continuamente actualizada en <ftp://ftp.ncbi.nlm.nih.gov/blast/db/FASTA/nr.gz>. La fecha de descarga de la base de datos fue el 17-02-18.

##### **4.1.2.3. Base de datos de COG y CDD**

La base de datos COG (Cluster of Orthologous Groups) es una herramienta útil para anotación funcional. Está disponible como un recurso de clasificación de proteínas curadas en el NCBI (<http://www.ncbi.nlm.nih.gov/COG/>) (Galperin *et al.*, 2014).

La base de datos CDD (Conserved Domain Database) es un recurso de anotación de proteínas. Se encuentra disponible como PSSMs (Position-specific score matrices) para la rápida identificación de dominios conservados en secuencias de proteínas vía RPS-BLAST (Marchler-Bauer *et al.*, 2014).

### **4.1.3. Softwares**

#### **4.1.3.1. Sistema Operativo: BIOLINUX: GNU/LINUX - UBUNTU.**

Linux es un sistema operativo libre desarrollado especialmente para programadores porque ofrece un control total sobre la computadora. Tiene una gran cantidad de herramientas de programación (lenguajes e interfaces), es estable y rápido. Todas estas características le permiten ser el sistema operativo más usado en bioinformática en especial para manejar datos genómicos. La mayoría de softwares desarrollados dentro del campo de la biología computacional o bioinformática son desarrollados para Linux.

BIOLINUX es una plataforma bioinformática desarrollada para Linux (Field *et al.*, 2006), trabaja dentro de la distribución Ubuntu 14.04. y contiene más de 250 paquetes bioinformáticos.

#### **4.1.3.2. Lenguajes de programación**

##### **4.1.3.2.1. Bash**

Bash es un lenguaje de programación a través del cual un usuario se comunica con el sistema operativo o una aplicación, en este caso el Shell de Linux.

##### **4.1.3.2.2. Python**

Python es un lenguaje de programación dinámico y legible. Representa una plataforma popular para todos los tipos de trabajo de bioinformática, que va desde el desarrollo de simples scripts hasta proyectos grandes y complejos. Actualmente, Python es uno de los lenguajes de programación más populares para los proyectos de secuenciación de nueva generación.

#### 4.1.3.2.3. R

R es un lenguaje de programación estadístico libre que es útil para los investigadores en el análisis de datos (R Core Team, 2013)

### 4.1.3.3. Programas

#### 4.1.3.3.1. MUSCLE

MUSCLE (Edgar, 2004) es un programa útil para realizar alineamiento múltiple de secuencias de ADN o proteínas.

#### 4.1.3.3.2. JModelTest2

JModelTest (Posada, 2008) es una herramienta útil para la selección estadística de los mejores modelos de sustitución nucleotídica. Implementa cinco estrategias diferentes de selección de modelos: hLRT, dLRT, criterio de información de Akaike y Bayesiano (AIC y BIC) y DT.

#### 4.1.3.3.3. Prokka

Prokka (Seemann, 2014) es un software para la rápida anotación de genomas procariotas y puede ser instalado en cualquier sistema Unix. A partir de una secuencia de ADN genómico pre-ensamblada en formato FASTA, Prokka sigue un pipeline para la anotación que consiste en los siguientes programas: Prodigal (Hyatt, 2010) para la predicción de secuencias codificantes (CDS), RNAmmer (Lagesen *et al.*, 2007) para la predicción de genes ARNr, Aragorn (Laslett y Canback, 2004) para genes ARNt, SignalP (Petersen *et al.*, 2011) para predicción de péptidos de señal, e Infernal (Nawrocki *et al.*, 2009) para la predicción de ARN no codificantes. Finalmente, Prokka produce 10 archivos de formatos diferentes en un directorio de salida.

#### 4.1.3.3.4. OrthoMCL

OrthoMCL (Li *et al.*, 2003) permite agrupar las secuencias ortólogas de todas las proteínas de cada genoma. A partir de las secuencias proteicas de los genomas anotados en Prokka, OrthoMCL primero realiza un *BLASTp all-against-all* para

definir los pares putativos de ortólogos o parálogos recientes basado en el BLAST recíproco. Los parálogos son identificados como genes dentro del mismo genoma que son recíprocamente más similares que cualquier secuencia de otro genoma. Luego OrthoMCL convierte los *p-values* del BLAST recíproco a una matriz de similitud normalizada que es analizado por Algoritmo de Agrupamiento de Markov (Markov Cluster algorithm-MCL). Finalmente, MCL produce un conjunto de clústeres con cada grupo que contiene un conjunto de ortólogos y/o parálogos.

#### 4.1.3.3.5. MySQL

OrthoMCL requiere una base de datos MySQL para ser configurado con el fin de cargar y procesar los agrupamientos.

#### 4.1.3.3.6. BLAST

Basic Local Alignment Search Tool (BLAST) (Altschul, 1990) encuentra regiones de similitud local entre secuencias. El programa compara las secuencias de nucleótidos o proteínas con bases de datos de secuencias y calcula la significación estadística de las coincidencias. BLAST se puede utilizar para inferir las relaciones funcionales y evolutivas entre las secuencias, así como ayudar a identificar a los miembros de las familias de genes.

#### 4.1.3.3.7. R 3.4.0 y RStudio

R (<https://www.r-project.org/>) es un software libre para análisis estadísticos y generación de gráficos. R compila y corre en una amplia variedad de plataformas UNIX y Windows. RStudio es una plataforma para escritura de scripts en lenguaje R con interfaz amigable. En el presente trabajo se usaron los siguientes paquetes:

*ggplot2*: Es un paquete de visualización y elaboración de gráficos en R (Wickham *et al.*, 2016).

*gplots*: Contiene muchas herramientas para generación de datos (Warnes *et al.*, 2016), entre ellas la función *heatmap.2*.

*Ape* y *Phangorn*: *APE* (Analysis of Phylogenetics and Evolution) (Paradis *et al.*, 2018) y *Phangorn* (Schliep *et al.*, 2018) son paquetes en R utilizados para reconstrucción filogenética.

*Vegan*: El paquete proporciona herramientas para análisis de diversidad, análisis de disimilaridad y ordenación de comunidad en ecología. Pero la mayoría de herramientas multivariadas pueden ser usadas para otros tipos de datos (Oksanen *et al.*, 2018).

*dendextend*: Es un paquete en R para la creación y comparación visual de diagramas de árboles. Proporciona funciones útiles para la manipulación de *objetos dendogramas* y opciones avanzadas para comparación de árboles filogenéticos de uno a otro (Galili, 2015).

*Kpax2*: Es un paquete que tiene el propósito de agrupamiento de grandes conjuntos de datos de tipo categórico. La salida de *k-pax2* consiste de una clasificación tanto de filas (unidades estadísticas) y columnas (variables estadísticas) que son proporcionados en una matriz de datos (Pessia *et al.*, 2015).

#### 4.1.3.3.8. Phipack

El software Phipack (Bruen, 2006) implementa algunas pruebas para evaluar la recombinación en una amplia variedad de circunstancias. Específicamente el “Pairwise Homoplasy Index”, Maximum X2 y el “Neighbour Similarity Score”. El programa Phi puede utilizarse para producir un *p-value* de recombinación dentro de un conjunto de datos.

#### 4.1.3.3.9. trimAL

trimAL (Capella-Gutierrez *et al.*, 2009) es una herramienta usada para la eliminación de regiones pobremente alineadas en un alineamiento incrementando la calidad de los análisis subsecuentes. La edición de un alineamiento es complicada en los análisis filogenéticos a larga escala que

utilizan miles de alineamientos. Por ello trimAL es adecuado para los análisis filogenómicos.

#### 4.1.3.3.10. FASconCAT

FASconCAT (Kück y Meusemann, 2010) es un script en lenguaje perl desarrollado para concatenar rápidamente diferentes tipos de secuencias en una supermatriz. A partir de un archivo de entrada en formato FASTA o PHYLIP genera un archivo de salida en los mismos formatos que pueden ser directamente cargado a un programa filogenético.

#### 4.1.3.3.11. ProtTest 3.3

ProtTest 3 (Darriba *et al.*, 2011) es una nueva versión para computación de alta performance (HPC) desarrollada para la plataforma UNIX usada para la selección del mejor modelo de sustitución de aminoácidos. ProtTest 3 hace esta selección encontrando el modelo en la lista de candidatos con el Criterio de Información de Akaike (AIC), el Criterio de Información Bayesiano (BIC) o el Criterio de Teoría de Decisión (DT).

#### 4.1.3.3.12. RAxML

RAxML (Stamatakis, 2014) es un programa popular para análisis filogenético de grandes conjuntos de datos bajo la metodología de *máximum likelihood*.

#### 4.1.3.3.13. RPS-BLAST

RPS-BLAST o PSI-BLAST inversa (Fryedberg *et al.*, 2000) es parte del paquete NCBI-BLAST. RPS-BLAST permite la comparación de una secuencia *query* contra una base de datos de *position-specific scoring matrices* (PSSMs) para la clasificación y anotación funcional de la secuencia.

### 4.1.4. Hardware

#### 4.1.4.1. Workstation

Procesador: Intel(R) Core(TM) i7-6500CPU 2.500GHz 2.6GHz



Memoria RAM: 16.00GB

Sistema: Sistema Operativo de 64 bits, procesador x64

## **4.2. Métodos**

### **4.2.1. Elección de secuencias genómicas de *Pasteurella multocida***

Las secuencias genómicas de *Pasteurella multocida* fueron descargados de la base de datos del sitio ftp del NCBI (<ftp://ftp.ncbi.nlm.nih.gov/genomes/>). Los genomas fueron seleccionados procurando tener una data heterogénea en términos de i) especificidad de hospedero (aves, cerdos, bovinos, conejos, etc.) ii) enfermedad (cólera aviar, coriza, rinitis atrófica, pasteurelosis neumónica y septicemia hemorrágica), y iii) calidad del ensamblaje (genomas que no tengan más de 100 contigs). Por tanto, la data genómica usada en este trabajo es una representación imparcial de la especie *Pasteurella multocida* que se encuentra disponible en la base de datos del Genbank.

### **4.2.2. Re-anotación de los genomas de *Pasteurella multocida***

Dado la heterogeneidad de métodos y pipelines usados en la anotación genómica desde el primer genoma secuenciado (cepa Pm70) en 2001 hasta la actualidad, y a fin de evitar sesgos técnicos en la comparación genómica posterior se sugiere actualizar la anotación (Koehorst *et al.*, 2016). Por tanto, se realizó una reanotación usando Prokka V1.11 a partir de las secuencias nucleotídicas de los genomas recuperados del NCBI en formato FASTA. Se usó Prodigal para la predicción de secuencias codificantes (CDS), RNAmmer para la predicción de genes ARNr y Aragorn para genes ARNt. Los genomas re-anotados en formato FASTA fueron asignados con nombres representativos que permita el posterior análisis.

### **4.2.3. Predicción de familias génicas ortólogas: Recuperación de genes ortólogos.**

Todas las secuencias proteicas re-anotadas en formato FASTA de las cepas de *P. multocida* usadas en este estudio se procesaron para la predicción de ortólogos usando OrthoMCL (Li *et al.*, 2003). Este programa primero realiza una

selección y eliminación de secuencias de baja calidad, luego un *all-against-all* BLASTp define agrupamientos putativos de ortólogos o parálogos recientes basado en los *hits* del BLAST recíproco usando *Markov Cluster Algorithm* (MCL). Para el caso, solo los hits del BLAST recíproco que muestran una identidad > 50% de identidad y 50% de cobertura con cada una fueron considerados como ortólogos. Los genes parálogos fueron eliminados de todas las familias génicas recuperadas de OrthoMCL. Un *bash script* fue usado para transformar la tabla de familias génicas en una matriz binaria (pan-matriz) donde las filas representen familias de ortólogos y las columnas representen los genomas. La matriz en formato '1', '0', es decir '1' representando para presencia y '0' para ausencia de los genes en el respectivo genoma, para cada familia de genes ortólogos.

#### 4.2.4. Perfil y tamaño del pangenoma de *Pasteurella multocida*

Las curvas de rarefacción y acumulación para las estimaciones de tamaño de genoma core y pangenoma se realizaron creando unos *scripts* en el paquete de software R a partir de la matriz de presencia/ausencia (pan-matriz) obtenida en el paso anterior. El número de genes core compartidos por todos los genomas, y el número acumulativo de diferentes genes fueron calculados aumentando de uno en uno todos los genomas. Los cálculos aleatorios se llevaron a cabo con 100 repeticiones, aleatorizando el orden de los genomas cada vez, para obtener el tamaño medio y el error estándar del core- y pan- genoma. Los tamaños del pangenoma y genoma core después de la adición de cada genoma fueron calculadas usando las siguientes fórmulas en R.

$$N_{pan} = \sum_{i=1}^n f_{pan}(G_i) \quad \text{y} \quad N_{core} = \sum_{i=1}^n f_{core}(G_i)$$

Donde  $G_i$  representa la familia génica  $i^{th}$ ,  $n$  es el número total de familias génicas distintas obtenidas del completo conjunto de datos y tamaños del core y pangenoma,  $(N_{pan}/N_{core})$  representa el tamaño de pan/genoma core después de la adición de  $n^{th}$  genoma del conjunto de datos.

$f_{pan}(G_i) = 1$ , si uno o más genomas entre los  $n$  genomas contiene al menos un miembro de la familia de genes  $G_i$ .

$f_{pan}(G_i) = 0$ , si ninguno de los  $n$  genomas contienen un miembro de la familia de genes  $G_i$ .

$f_{core}(G_i) = 1$ , si todos de los  $n$  genomas contienen a menos un miembro de  $G_i$ .

$f_{core}(G_i) = 0$ , si al menos uno de los  $n$  genomas no contiene cualquier miembro de  $G_i$ .

Como reportó Tettellin *et al.* en 2008 la ley de Heaps (Heaps 1987), una ley empírica originalmente empleada en el campo de la recuperación de información, puede ser usada para describir el tamaño de pangenoma bacteriano. La ecuación de esta ley es:

$$N_{pan} = kx^\gamma$$

Donde  $N_{pan}$  es el tamaño del pangenoma para un número “ $x$ ” de genomas calculados,  $k$  y  $\gamma$  son coeficientes calculados en el ajuste de la curva. Finalmente,  $\alpha = (\gamma - 1)$ ; donde, si  $\alpha < 1$ , el pangenoma es “abierto”, y si  $\alpha > 1$ , el pangenoma es cerrado.

#### 4.2.5. Anotación funcional: Análisis de agrupamiento de grupos ortólogos (COG) de proteínas

El conjunto de genes core, accesorios y únicos del pangenoma de *P. multocida* fueron asignados a 20 diferentes categorías COG. Los COG fueron anotados mediante RPSBLAST de cada secuencia proteica contra la base de datos CDD COG del NCBI con un e-value *cutoff* < 0.001.

#### 4.2.6. Análisis Filogenético

Se construyeron árboles filogenéticos basados en secuencias de ADN o proteínas y árboles filogenéticos basados en contenido génico para entender la diversidad genómica de las cepas bajo estudio.

##### 4.2.6.1. Filogenias basado en secuencia

###### 4.2.6.1.1. Filogenia MLST

La Tipificación de Secuencia de Multilocus (MLST) *in silico* fue realizado para determinar los tipos de secuencia (STs) y estructuras de complejo clonal (CC) de cada cepa (genoma elegido) según el esquema MLST RIRDC del PubMLST (<https://pubmlst.org/pmultocida/>) usando los genes *housekeeping adk*, *est*, *pmi*,

*zwf*, *mdh*, *pgi* y *gdh* recuperados de cada genoma usado en este estudio. Primero se realizó el alineamiento múltiple de cada secuencia con MUSCLE, luego las 7 secuencias fueron concatenadas con FASconCAT para construir finalmente un árbol filogenético usando *maximum likelihood* en RAxML con 1000 de *bootstrap* luego de calcular el mejor modelo de sustitución con JModelTest.

#### 4.2.6.1.2. Análisis filogenómico del genoma core

Para realizar la filogenia del genoma core, los genes core (familias de ortólogos que en la matriz de presencia/ausencia están presente en todos los genomas, es decir todos '1') fueron extraídos en secuencias nucleotídicas y proteicas en formato FASTA usando un *bash script*. Todas las secuencias ortólogas de todas las familias fueron renombrados con el nombre del genoma al que pertenecen para poder crear una super-matriz mediante un *bash script*. El alineamiento de todas las secuencias se llevó a cabo con MUSCLE en una combinación con comandos *bash*. Los alineamientos de aminoácidos de estos genes fueron editados (*trimming*) con trimAL, para que todas las secuencias alineadas posean la misma longitud. La salida de los archivos editados con trimAL en formato *PHYLIP* fueron usados para ser concatenados y generar una super-matriz usando FASconCAT *perl script* (Kuck y Meusemann 2010). Esta supermatriz conteniendo todos los genes core concatenados fue sometida a ProtTest 3 para evaluar el modelo evolutivo de sustitución de aminoácidos. Finalmente, la construcción del árbol filogenético se realizó usando *maximum likelihood* en RAxML con 100 réplicas de *bootstrap*.

#### 4.2.6.1.3. Análisis de recombinación

Debido a que parte de los genes *housekeeping* del genoma core están afectados por recombinación homóloga y que esto puede afectar la reconstrucción filogenética, todas las secuencias fueron evaluadas para recombinación con el paquete PhyPack usando un *script* en *python*, aquellos genes que muestran evidencia de recombinación fueron excluidos y con los genes restantes se construyó un nuevo árbol filogenético denominado 'genoma core mínimo' con las mismas herramientas que el paso anterior.

#### **4.2.6.2. Filogenia basada en contenido génico (árbol pangenómico)**

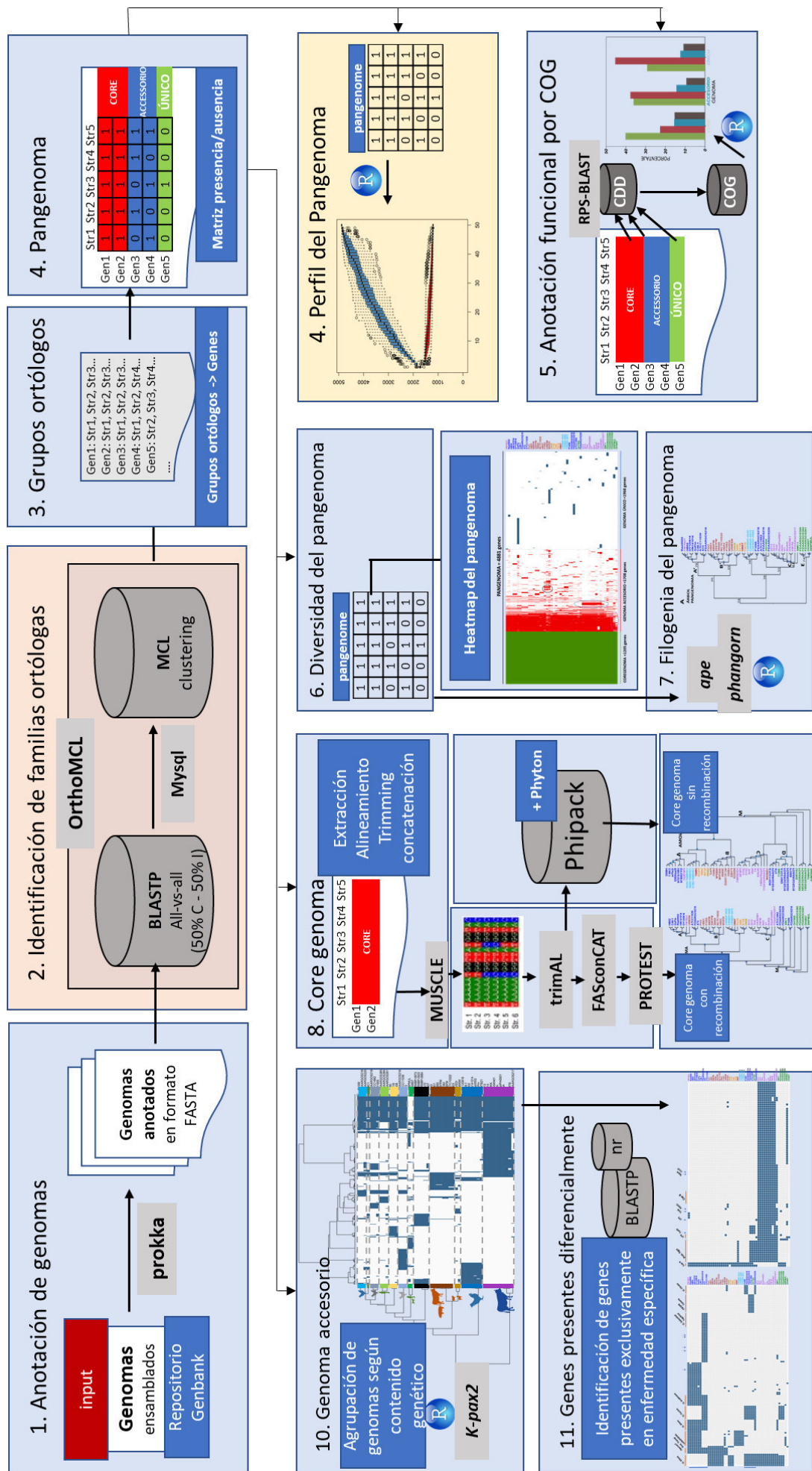
La matriz binaria de presencia/ausencia de genes ortólogos '1','0' (pan-matriz) fue usada para construir un árbol pan-genómico usando los paquetes *Ape* y *phangorn* en R, para el cálculo de distancia se utilizó el algoritmo distancia de Jaccard. Finalmente, el árbol fue construido usando *Neighbour-Joining* del paquete *Ape* usando 1000 réplicas de *bootstrap*. Finalmente, los árboles contruidos a partir del genoma core y pangenoma fueron usados para construir un tanglegrama, es decir compararlos frente a frente para evaluar la congruencia entre ambas reconstrucciones.

#### **4.2.7. Análisis de la diversidad de genoma accesorio**

La visualización es una máxima del análisis de datos masivos, por ello se utilizaron herramientas en R para la elaboración de *heatmaps* para visualizar la diversidad del genoma accesorio y único. Puesto que el genoma accesorio es considerado a tener genes con capacidad de adaptación a nicho, se utilizó un enfoque bayesiano con K-pax2 para predecir el agrupamiento de genes del genoma accesorio y reconocer aquellos conjuntos de genes que son responsables de esa agrupación. Posteriormente se realizó la caracterización de los genes mediante *BLASTp* contra la base de datos de proteínas no redundantes (nr) del NCBI. Los nombres de las proteínas predichas fueron recuperados usando *blastcmd*, para luego ser visualizadas en *heatmaps* según su asociación a enfermedad.

#### **4.2.8. Pipeline**

La figura 3 muestra el pipeline utilizado en el presente trabajo. Se muestra el flujo de trabajo bioinformático desde la obtención de genomas hasta la generación del pangenoma, construcción de árboles filogenéticos y *heatmaps*.



**Figura 3. Pipeline de la metodología usada en la presente investigación.** Se grafica el flujo de trabajo bioinformático desde la obtención de genomas hasta la generación del pangenoma, construcción de árboles filogenéticos y *heatmaps*

## 5. RESULTADOS

### 5.1. Genomas seleccionados para el análisis pangenómico

De un total de 133 genomas de *Pasteurella multocida* disponibles en la base de datos pública del GenBank (marzo 2017) se llegó a seleccionar 50 secuencias genómicas de los cuales 23 son completos y 27 borradores con menos de 100 contigs. El tamaño de los genomas seleccionados está dentro del rango de 2.08-2.54 Mb y el porcentaje de G + C entre 36.9-40.9%. Las secuencias genómicas pertenecen a cepas aisladas de distintos hospederos y enfermedades (Tabla 1). 14 genomas son cepas de aves con cólera aviar (aislados de pavos, patos y pollos), 11 son genomas de cepas provenientes de animales con pasteurelosis neumónica (aislados de bovinos, cerdos, ovino y alpaca), 11 genomas de cepas de animales con septicemia hemorrágica (aislados de bovinos y búfalos), 7 genomas de cepas asociadas a infecciones en humanos (aislados de humanos y perros), 3 genomas de cepas de conejos con coriza, un genoma de cepa de cerdo con rinitis atrófica (cepa HN06), 3 genomas de cepas avirulentas (Pm70 de ave, 3840 de cerdo y 617/19 de bovino).

**Tabla 1: Lista de genomas de *P. multocida* usadas en el presente estudio distribuidos por hospedero y enfermedad**

Enfermedad	Hospedero (Cantidad)	Total de cepas
Cólera aviar	Pollos (5) Patos (3) Pavos (5)	14
Pasteurelosis neumónica	Bovino (6) Cerdos (3) Ovino (1) Alpaca (1)	11
Septicemia hemorrágica	Bovino (4) Bisonte (1) Búfalo (6)	11
Coriza	Conejo (3)	3
Avirulentos	Pavo (1) Cerdo (1) Bovino (1)	3
Rinitis atrófica	Cerdo (1)	1
Pasteurelosis como Zoonosis	Perro (2) Humano (5)	7

\*ND: Especie aviar no determinada

En la tabla 2 se resumen las características básicas de los genomas de *Pasteurella multocida*, 36 de los aislados han sido caracterizados genéticamente y serotípicamente, 14 de ellos solo son secuencias genómicas disponibles sin serotipificación experimental.

## **5.2. La re-anotación de genomas muestra una diversidad en contenido genético.**

Los 50 genomas fueron descargados en formato Fasta (.fna) y cargados en el programa Prokka para su anotación automática usando prodigal como algoritmo de predicción de secuencias proteicas, como resultado el número de secuencias codificantes de proteínas predichas para cada uno de los 50 genomas están entre 1930 y 2923, con un promedio de 2200 secuencias codificantes por genoma. Esta variabilidad del repertorio genético en cuanto a número de proteínas por cepa sugiere una diversidad importante en el genoma de este organismo (tabla 2). El total de secuencias codificantes anotados para los 50 genomas es de 109832.



**Tabla 2. Características de los genomas de las 50 cepas de *Pasteurella multocida* usadas en este estudio.**

N°	Nombre de cepa	Hospedero	Enfermedad	Tamaño (Mb)	GC%	Número de CDS predicho con Prokka	Serogrupo/ Serotipo	Bioproject Acceso	Completo (C)/ Borrador (# de contigs)
1	Pm70	Pavo	Avirulento	2.26	40.4	2028	F:3	PRJNA39	C
2	3480	Cerdo	Avirulento	2.38	40.3	2213	A	PRJNA32177	C
3	671/90	Bovino	Avirulento	2.25	40.2	2048	A:3	PRJNA188485	21
4	HN141014	Pato	CA	2.28	40.3	2095	A	PRJNA303094	16
5	Razi0002	Aviar	CA	2.29	40.4	2193	A:1	PRJNA360234	C
6	HB02	Pato	CA	2.21	40.3	2019	A	PRJNA320086	46
7	ATCC 1662	Pavo	CA	2.27	40.3	2057	A:4	PRJNA300371	15
8	FDAARGOS_218	Pavo	CA	2.34	40.4	2137	A:3	PRJNA231221	C
9	FDAARGOS_219	Pavo	CA	2.51	40.4	2391	A:3	PRJNA231221	13
10	FDAARGOS_220	Pavo	CA	2.39	40.4	2189	A:5	PRJNA231221	2
11	P1059	Pavo	CA	2.31	40.3	2119	A:3	PRJNA152827	C
12	ATCC 11039	Pollo	CA	2.27	40.3	2082	A:1	PRJNA300368	18
13	C48-1	Pollo	CA	2.26	40.2	2090	A:1	PRJNA298956	17
14	GX-Pm	Pollo	CA	2.29	40.3	2902	A	PRJNA277615	8
15	X73	Pollo	CA	2.27	40.3	2078	A:1	PRJNA155371	C
16	DY120818	Pato	CA	2.28	40.3	2096	A:1	PRJNA302563	15
17	FDAARGOS_216	Pollo	CA	2.30	40.4	2092	A:1	PRJNA231221	C
18	CIRMBP-0873	Conejo	Pasteurelosis	2.48	40.4	2392	A	PRJNA355236	C
19	CIRMBP-0884	Conejo	Pasteurelosis	2.46	40.5	2390	A	PRJNA355236	C
20	CIRMBP-0922	Conejo	Pasteurelosis	2.28	40.2	2120	A	PRJNA355236	14
21	HN06	Cerdo	R.A.	2.41	40.2	2252	D	PRJNA80883	C
22	NCTC10322	Cerdo	Neumonía	2.27	40.4	2076	A	PRJEB6403	C

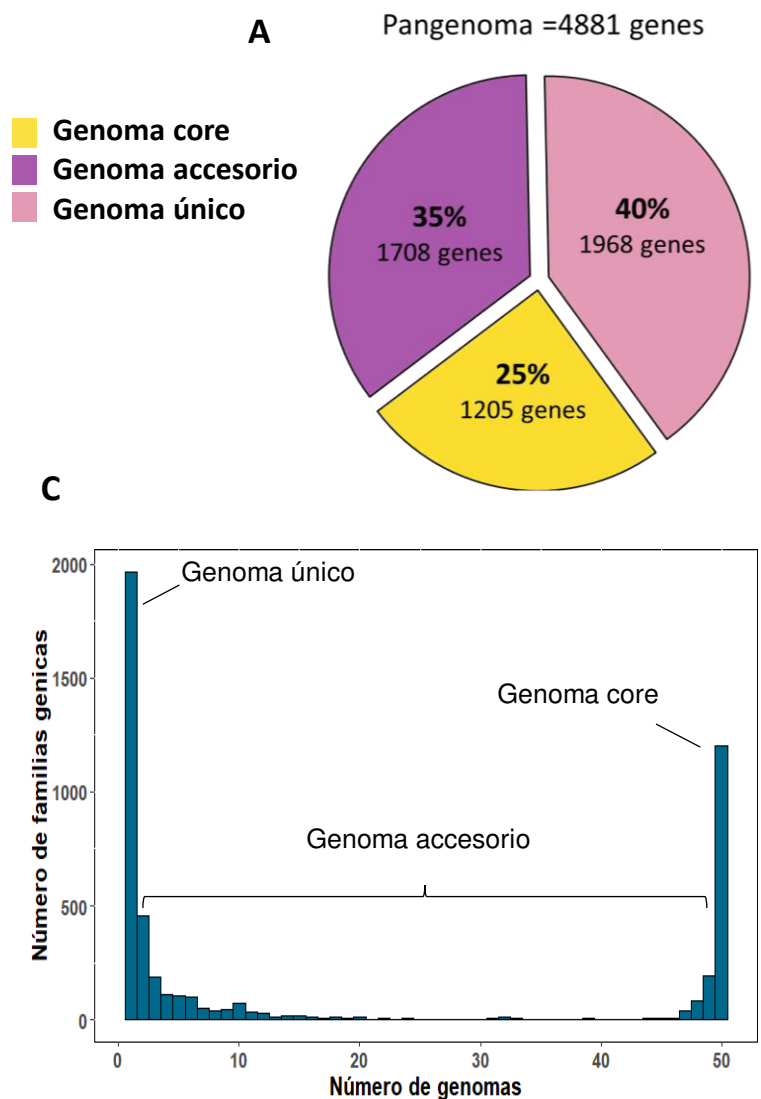
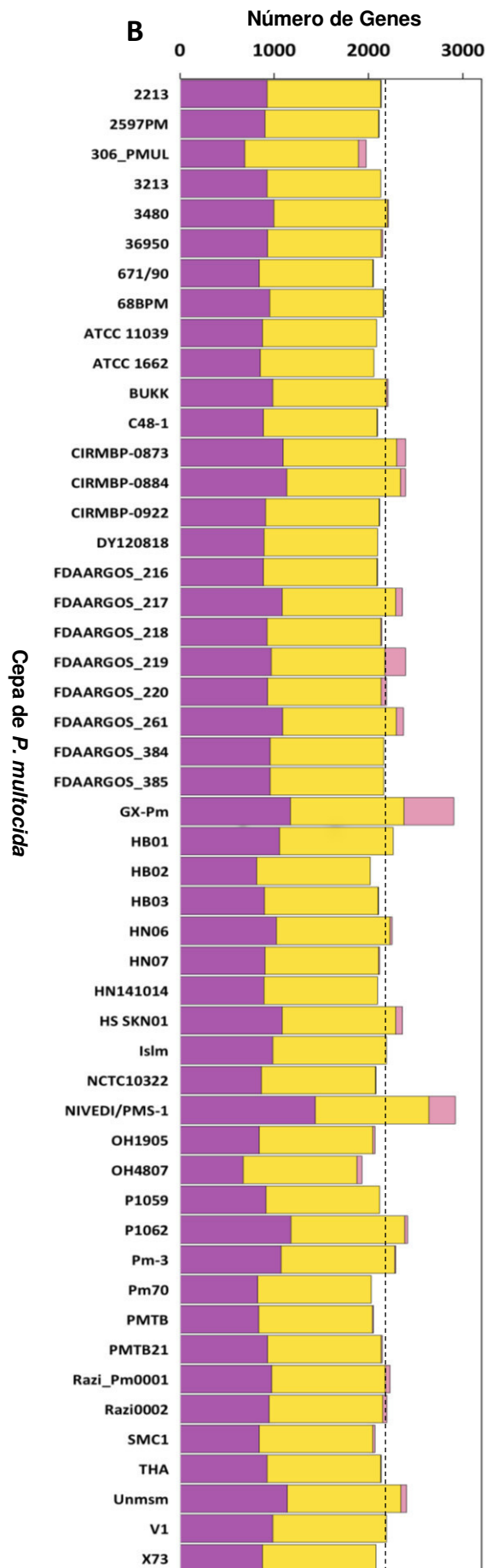
23	2597PM	Bovino	Neumonía	2.30	40.3	2111	A	PRJNA306895	30
24	68BPM	Bovino	Neumonía	2.30	40.2	2161	A	PRJNA306895	26
25	UNMSM	Alpaca	Neumonía	2.44	40.2	2401	A:6	PRJNA276484	82
26	36950	Bovino	Neumonía	2.35	40.4	2152	A:3	PRJNA70829	C
27	HB01	Bovino	Neumonía	2.42	40.3	2260	A	PRJNA210009	C
28	HB03	Cerdo	Neumonía	2.31	40.4	2105	A	PRJNA83131	C
29	P1062	Bovino	Neumonía	2.70	40.7	2412	A:3	PRJNA201507	C
30	Pm-3	Bovino	Neumonía	2.42	40.3	2289	A	PRJNA312874	C
31	HN07	Cerdo	Neumonía	2.33	40.3	2117	F	PRJNA231703	C
32	NIVEDI/PMS-1	Oveja	Neumonía	2.54	40.2	2923	A	PRJNA339918	67
33	FDAARGOS_217	Bisonte	SH	2.44	40.5	2362	B:2	PRJNA231221	7
34	2213	Bovino	SH	2.30	40.4	2128	B:2	PRJNA251431	33
35	3213	Bovino	SH	2.30	40.4	2127	B:2	PRJNA251433	31
36	BUKK	Bovino	SH	2.36	40.4	2206	B:2	PRJNA244049	37
37	Razi_Pm0001	Bovino	SH	2.36	40.4	2230	B:2	PRJNA352138	C
38	HS SKN01	Bufalo	SH	2.45	40.7	2361	B:2	PRJNA339731	48
39	Islm	Bufalo	SH	2.35	40.3	2187	B:2	PRJNA244093	32
40	PMTB	Bufalo	SH	2.20	40.4	2049	B:2	PRJNA172191	8
41	PMTB2.1	Bufalo	SH	2.32	40.3	2143	A	PRJNA236642	C
42	THA	Bufalo	SH	2.30	40.4	2131	B:2	PRJNA244069	29
43	V1	Bufalo	SH	2.35	40.3	2187	B:2	PRJNA244097	27
44	OH1905	Perro	Zoonóticos	2.27	40.4	2068	A	PRJNA192965	C
45	OH4807	Perro	Zoonóticos	2.08	39.7	1930	N.D	PRJNA192964	C
46	FDAARGOS_261	Humano	Zoonóticos	2.47	40.3	2369	A	PRJNA231221	3
47	SMC1	Humano	Zoonóticos	2.27	40.3	2070	A	PRJNA301781	23
48	306_PMUL	Humano	Zoonóticos	2.17	36.9	1927	N.D	PRJNA267549	30
49	FDAARGOS_384	Humano	Zoonóticos	2.35	40.5	2161	A	PRJNA231221	C
50	FDAARGOS_385	Humano	Zoonóticos	2.35	40.5	2161	A	PRJNA231221	C

CA= Cólera aviar, SH= Septicemia hemorrágica, RA= Rinitis atrófica, N.D = No determinado

### 5.3. Agrupación de familias génicas y características generales del pan-genoma

Las 109832 secuencias codificantes de los 50 genomas de *P. multocida* fueron agrupadas por OrthoMCL (identidad de 50% y *e-value* de  $10^{-5}$ ) dando como resultado un pan-genoma con 4881 familias ortólogas (genes en adelante), alrededor de 2.2 veces el número promedio de secuencias codificantes por genoma (2200 genes). De las 4881 familias de genes solo 1205 (24.69%) son compartidos por todas las cepas, constituyendo el genoma core de *P. multocida*; en tanto 1708 (34.99%) genes del pangenoma son genes accesorios (es decir, genes presentes en dos o más genomas, pero no en todos) y los restantes 1968 (40.31%) genes han sido identificados como genes únicos (Figura 4a). Hurtado *et al.*, (2018) reportan un tamaño de pangenoma de *P. multocida* de 3585 genes con 23 genomas analizados, mostrando diferencias en contenido genético y presencia de genes únicos en algunas cepas en concordancia con un alto nivel de diversidad genética en esta bacteria mostrada en este pangenoma. En promedio el genoma de *P. multocida* codifica 2200 genes, por tanto, el tamaño del genoma core representa poco más de la mitad (55%) de cada genoma de *P. multocida* (Figura 4b). Los genomas Gx-Pm (aviar) y NIVEDI (ovino) son las que poseen una mayor cantidad de genes accesorios y únicos, el primero fue objeto de estudio por Yu *et al.* (2016) en el que la presencia de 5 islas genómicas, 4 IS (secuencias de inserción) y 2 profagos suman una cantidad de proteínas que es mucho mayor al de un genoma promedio de *P. multocida*.

La figura 4c muestra la distribución de las familias de genes según el número de genomas de *P. multocida*, En esta, las tres categorías del pangenoma (genoma core, accesorio y único) se logran distinguir. La distribución en forma de U indica que la mayoría de genes en *P. multocida* están ya sea compartido por muy pocos genomas o son encontrados en casi todos los genomas.



**Figura 4: Características del Pangenoma de *P. multocida*.**

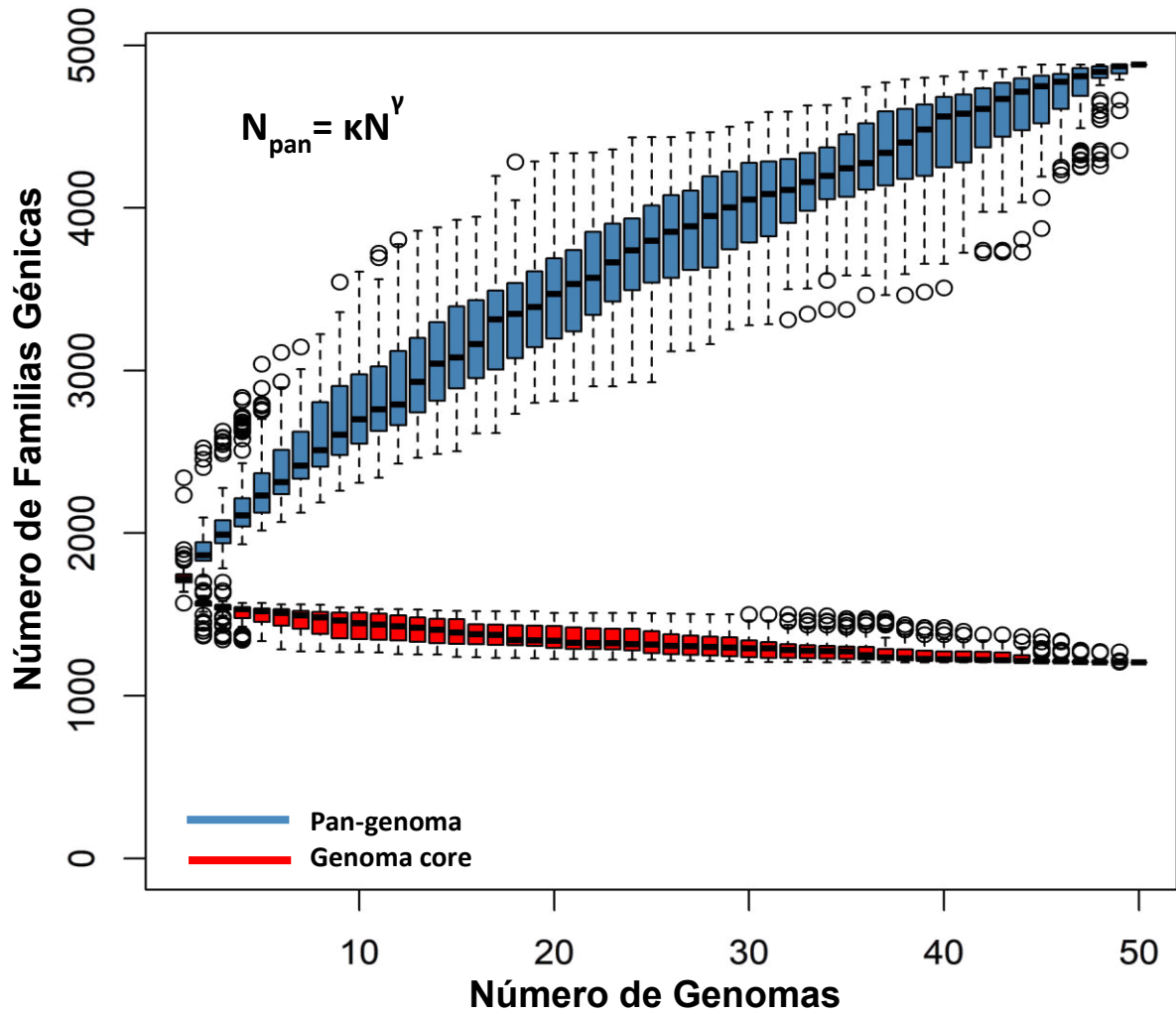
**A)** El pangenoma de *P. multocida* se divide en genoma core (compartido por todas las cepas), genoma accesorio (compartido por dos o más cepas, pero no por todas), genoma único (presente solo en cepas únicas). **B)** Distribución de los genes en el pangenoma de por cada cepa de *P. multocida*, las líneas punteadas representan el genoma promedio de *P. multocida* (2200 genes). **C)** Histograma del número de las familias génicas por número de genomas. La primera columna representa el número de genes presentes en un solo genoma (genoma único), en tanto la última, el número de genes presentes en 50 genomas (genoma core).

#### 5.4. Plot de pangenoma y genoma core

La figura 5 muestra la expansión gradual del pan-genoma y la contracción del genoma core de *P. multocida* según la adición secuencial de genomas. Cualquier posible sesgo debido a la adición de genomas fue eliminado mediante permutaciones aleatorias en el orden de adición de cada genoma. El tamaño del pangenoma va incrementándose según se aumenta la cantidad de genomas hasta alcanzar 4881 familias génicas cuando 50 genomas de *P. multocida* fueron añadidos (gráfico de cajas azul en la figura 5). La extrapolación del pangenoma usando el modelo de regresión “power-law” de la Ley de Heaps sugiere que *P. multocida* tiene un pan-genoma “abierto” expandiéndose a una tasa moderada con  $\alpha = 0.73$ . Esto significa que teóricamente la expansión del pangenoma de *P. multocida* no tiene límite.

Por otro lado, el genoma core parece haber llegado a un estado “cerrado” en el cual la adición de nuevas secuencias no disminuirá significativamente el número de genes core (gráfico de cajas roja en la figura 5).

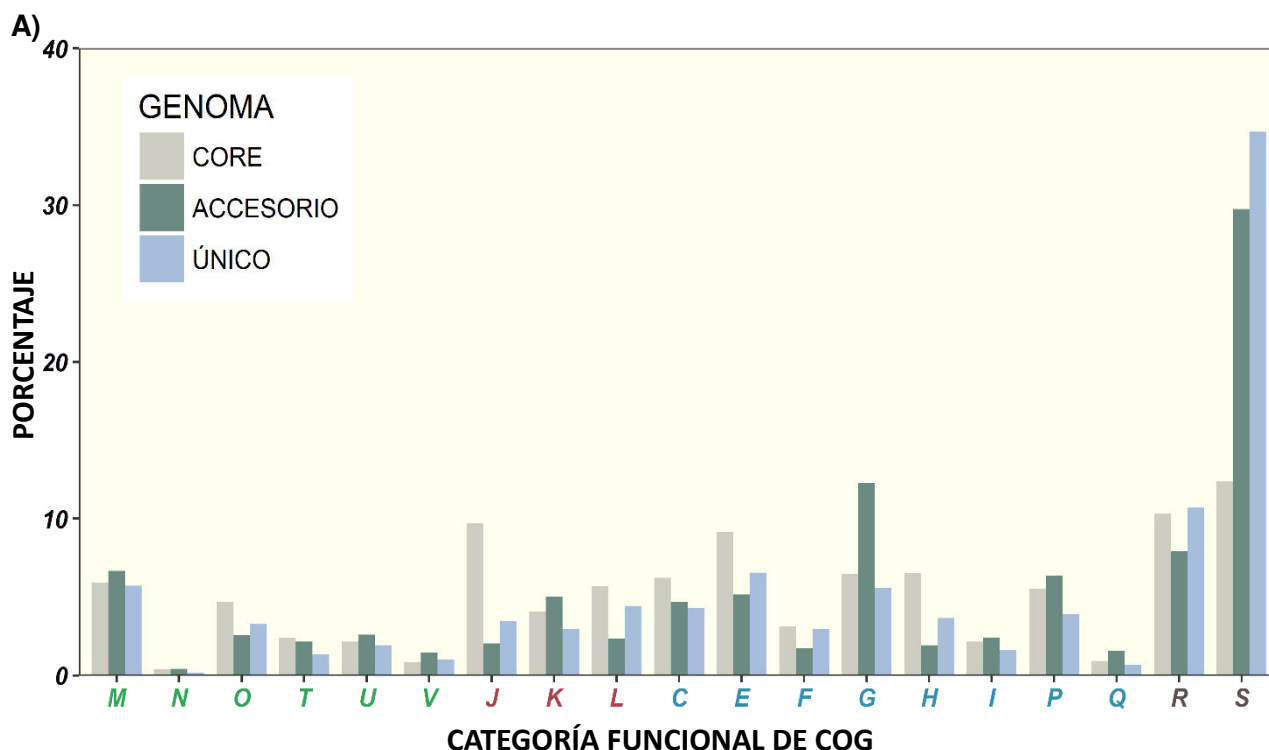
En la figura 5, los bordes inferior y superior de las cajas representan el primer cuartil (25avo percentil de los datos) y el tercer cuartil (75avo percentil de los datos), respectivamente. La línea central horizontal dentro de la caja indica la mediana de la muestra (50avo percentil) y los bigotes se extienden hasta 1.5 veces el rango intercuartílico. Los círculos presentes fuera de los bigotes son valores atípicos.



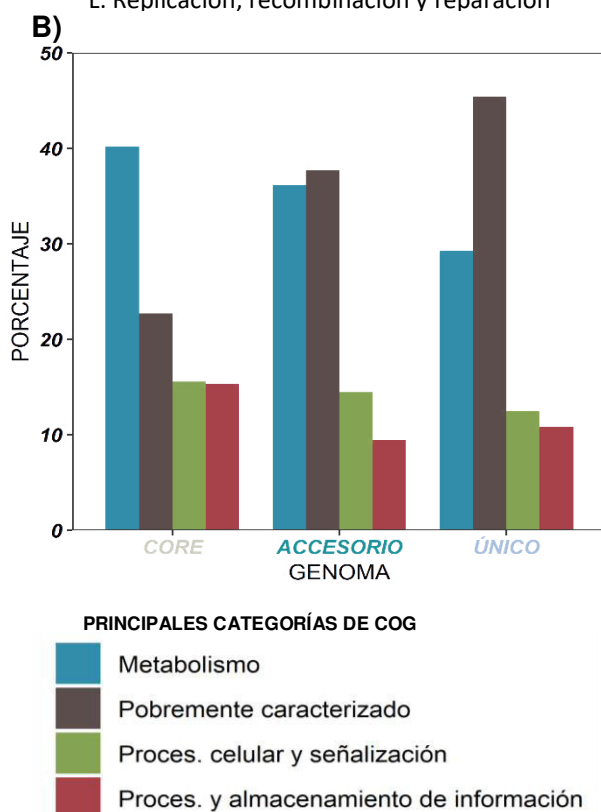
**Figura 5: Perfil del pangenoma y genoma core de *P. multocida*.** Las curvas de genes compartidos o genoma core (rojo) y del total de genes o pan-genoma (azul) para 50 cepas de *P. multocida* son ploteadas según el número de genomas incorporados. La secuencial adición de los genomas fue aleatorizada con 1000 permutaciones y estos valores son representados en cajas usando el software R. Para la ecuación del pangenoma  $N_{\text{pan}}$  es el número de genes esperados para un número (N) de genoma dado,  $\kappa$  y  $\gamma$  son parámetros libres definidos para ajustar la curva de expansión. El valor determinado para  $\gamma$  fue 0.27. Usando la Ley de Heap ( $\alpha = 1 - \gamma$ ), el valor de  $\alpha$  es 0.73, el cual indica un pangenoma abierto para *P. multocida*

## 5.5. Anotación funcional del pangenoma de *Pasteurella multocida*

La anotación funcional de genes en el pangenoma de *P. multocida* usando la base de datos COG revela una distribución única de categorías funcionales entre los distintos conjuntos del pangenoma. La fracción más grande del genoma core pertenece a genes clasificados como metabólicos (40%). El genoma core está asociado a procesos *housekeeping*, estos incluyen funciones no metabólicas tales como transcripción (4%), traducción, estructura y biogénesis ribosomal (10%). Por otro lado, genes involucrados en funciones metabólicas son también altamente conservados. Tales funciones incluyen metabolismo y transporte de aminoácidos (9%), transporte y metabolismo de carbohidratos (6%), transporte y metabolismo de coenzimas (7%). Juntos, los genes involucrados en procesos metabólicos, transcripción y traducción es 54% del genoma core de *P. multocida*. Por otro lado, una menor fracción de genes metabólicos es observada en los genomas accesorio y único (36 % y 29 % respectivamente). En estas últimas, la principal fracción de genes pertenece a categoría de función desconocida y pobremente caracterizado (37% y 45% para el genoma accesorio y único, respectivamente). El genoma accesorio tiene una amplia porción de genes con función desconocida o pobremente caracterizados (37%). Las funciones metabólicas involucradas en el genoma accesorio están representadas por metabolismo y transporte de carbohidratos (12%), metabolismo y transporte de aminoácidos (6%) y transporte de iones inorgánicos (6%). Es importante resaltar la presencia de genes relacionados al metabolismo y transporte de carbohidratos representados en el genoma accesorio en una proporción relativamente mayor que en el genoma core. El genoma único está representado principalmente por genes que codifican proteínas de función desconocida, secuencias de profagos y otras categorías pobremente caracterizadas (45%). Esta alta proporción de elementos móviles en el genoma único encontrado en *P. multocida* también se observa en otras bacterias como *Escherichia coli* y *Staphylococcus aureus* e indica que eventos de transferencia genética horizontal tiene un efecto importante en la evolución de *P. multocida* (Figura 6).



C: Producción y conversión de energía  
 E: Metabolismo y transporte de aminoácidos  
 F: Metabolismo y transporte de nucleótidos  
 G: Metabolismo y transporte de carbohidratos  
 H: Metabolismo y transporte de coenzima  
 I: Metabolismo y transporte de lípidos  
 J: Traducción, estructura ribosomal y biogénesis  
 K: Transcripción  
 L: Replicación, recombinación y reparación  
 M: Biogénesis de pared, membrana celular  
 N: Motilidad celular  
 O: Modificaciones postraduccionales  
 P: Metabolismo y transporte de iones inorgánicos  
 Q: Catabolismo, transporte de metabolitos secundario  
 R: Predicción de función única  
 S: Función desconocida  
 T: Mecanismo de traducción de señal  
 V: Mecanismos de defensa  
 U: Tráfico intracelular, secreción y transporte vesicular



**Figura 6: Anotación funcional de los genomas core, accesorio y único de *P. multocida*:**

**A)** 19 categorías funcionales de COG (*Cluster of orthologous Groups*) son asignados a las familias génicas de cada genoma (core, accesorio y único) y se muestran en porcentaje. Las letras representan las categorías funcionales y los colores a los principales grupos COG que se muestra en B.

**B)** Se representa el porcentaje de las 4 principales categorías funcionales COG presentes en los genomas core, accesorio y único: metabolismo, procesamiento y señalización, procesamiento y almacenamiento de información y pobremente caracterizados.

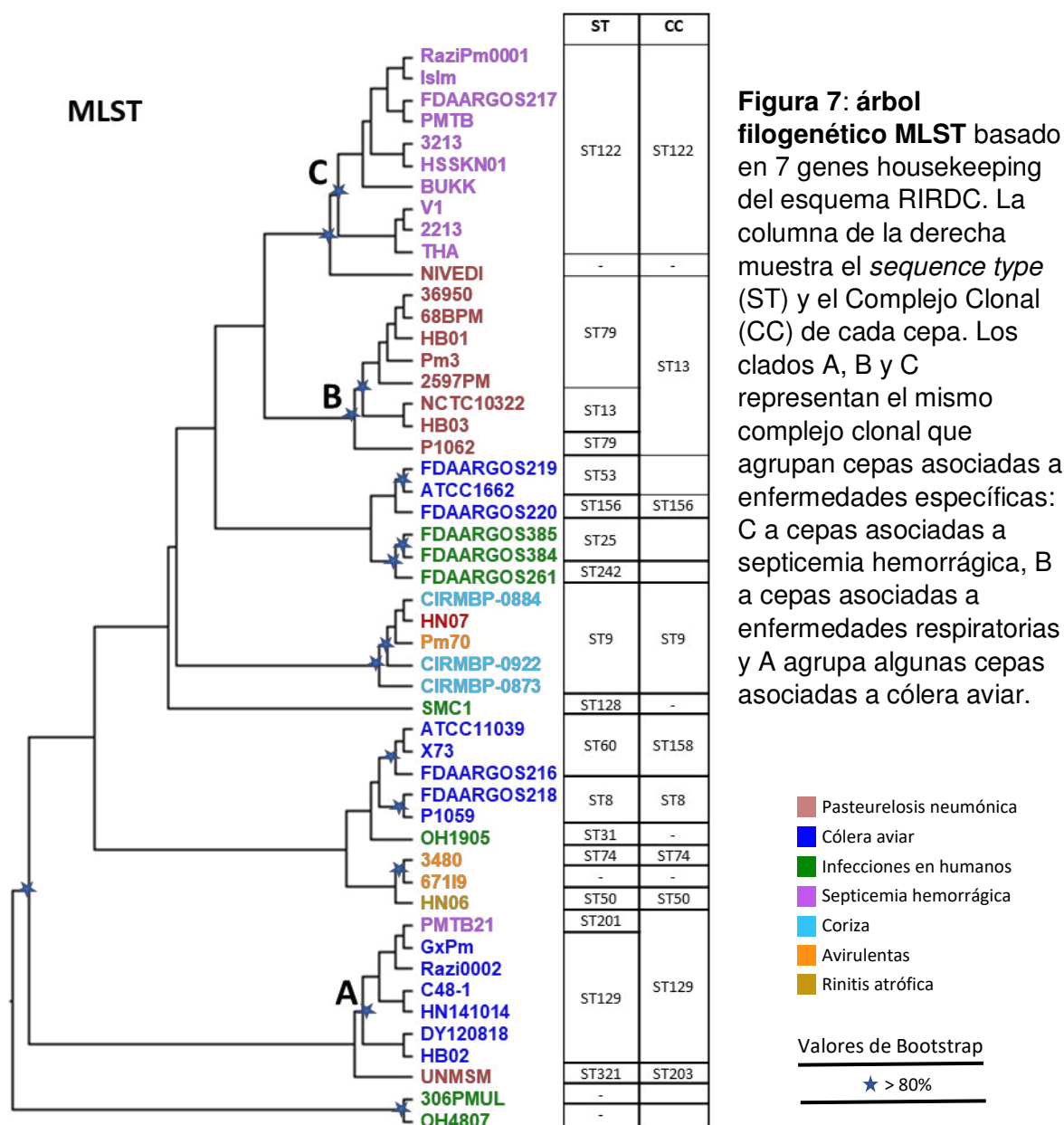


## 5.6. Análisis Filogenómico

Para revelar las relaciones evolutivas entre las cepas analizadas de *P. multocida*, se construyeron 3 árboles filogenéticos basado en MLST (*Multilocus Sequence Typing*), SNPs del genoma core y matriz de presencia ausencia del pan-genoma.

### 5.6.1. Análisis filogenético a partir de MLST *in silico*

Se construyó un árbol filogenético a partir de 7 genes *housekeeping* (*adh*, *est*, *pmi*, *zwf*, *mdh*, *gdh*, *pgi*) del esquema RIRDC identificados *in-silico* a partir del genoma de cada aislado usando *maximum likelihood* en el programa RAxML y el árbol se visualiza enraizado a punto medio (Figura 7). Cada genoma fue sometido a la base de datos del PubMLST del esquema RIRDC para recuperar los alelos de cada cepa. En el árbol se observa cierto agrupamiento por cepas asociados a enfermedad específica: grupos soportados por fuertes *bootstraps* (>80% (estrellas azules)) son observados en cepas relacionadas a Septicemia Hemorrágica (ST122, Clado C), 8 cepas relacionadas con enfermedades respiratorias también conforman un solo clado (Complejo Clonal (CC)-ST13, clado B) y el clado A muestra cepas asociadas a cólera aviar (CC-ST129). MLST utiliza alelos como unidad de comparación designados en *sequence types* (ST). Para los 50 genomas de este estudio, la designación de STs fue: para las 14 cepas de cólera aviar un total de 5 genotipos (ST129, ST60, ST53, ST156 y ST8). Para las 11 cepas asociadas a septicemia hemorrágica, las 10 cepas tipo capsular B fueron del mismo ST (ST122), solo la cepa PMTB21 es de ST201. En tanto, para las cepas asociadas a enfermedades respiratorias 4 genotipos fueron asignados (ST79, ST13, ST80 (estas tres del CC13), y ST321(cepa UNMSM)), pero para el genoma de NIVEDI no se encontró ningún ST en la base de datos. Finalmente, para las cepas asociadas a enfermedades de humanos, 4 genotipos fueron asignados (ST31, ST25, ST242 y ST128), con las cepas 306PMUL y OH4807 sin identificación de STs en la base de datos. Si bien el MLST es una técnica ampliamente usada para estudiar la diversidad bacteriana e identificar linajes que pueden ser patogénicas o no, el pequeño número de genes (siete) puede no ser representativo del genoma completo (Didelot y Maiden, 2010). Por ello a continuación se realizó un análisis filogenético del genoma core.



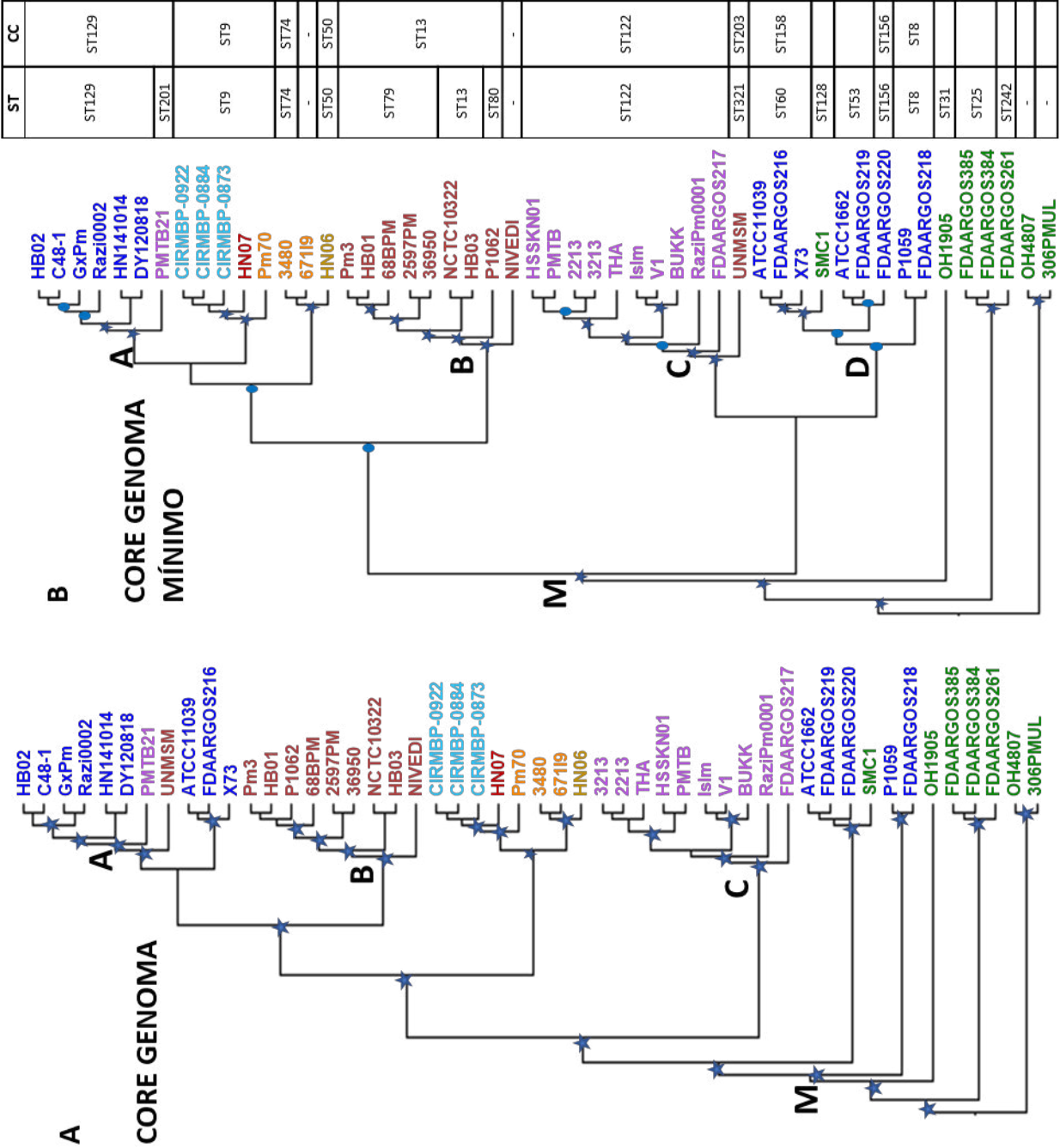
### 5.6.2. Análisis filogenético del Genoma core

Las secuencias de aminoácidos de las 1205 secuencias codificantes de copia única que forman parte del genoma core, fueron extraídas, alineadas y concatenadas resultando en una longitud de 304,798 aa. Este alineamiento sirvió para construir un árbol filogenómico usando RAxML con PROTGAMMAIJTTF determinado por PROTEST 3.0 como mejor modelo evolutivo. El árbol resultante se muestra en la figura 8A.

El impacto de la recombinación homóloga en la evolución bacteriana es importante y cuando ocurre en genes core genera desafíos para la reconstrucción filogenética. Por ello se utilizó el programa Phipack y un script escrito en Python para detectar genes que están afectados por recombinación. Un valor de  $p\text{-value} < 0.05$  se utilizó para considerar un gen con evidencia de recombinación. Con esta metodología se logró detectar 519 (42%) genes con evidencia de recombinación y los restantes 686 (58%) genes sin evidencia de recombinación fueron considerados como genoma core “mínimo”. Estos genes fueron alineados y concatenados en su secuencia de aminoácidos (total: 143,941 aa de longitud) y fue usado para construir un segundo árbol filogenético (Figura 8B).

Los árboles filogenómicos contruidos a partir del genoma core y el genoma core mínimo conservan los 3 clados (A, B y C) observados en el árbol contruido a partir del esquema MLST, pero con distinta relación entre estos. Ambos árboles filogenómicos muestran que las cepas asociadas a infecciones en humanos son las más distantes (6 de los 7, conformados en clados parafiléticos), separadas del resto de cepas patogénicas para otros animales. En la filogenia del genoma core mínimo las cepas asociadas a cólera aviar están distribuidas en dos clados distintos (A y D), en tanto las cepas asociadas a coriza en conejos y las dos cepas de serotipo F (Pm70, HN07) conforman un grupo que también está presente en el árbol MLST y genoma core. Otro grupo conservado en los 3 árboles es la que conforman (HN06, 3480, 671119), estas cepas tienen un denominador común: la cepa 3480 es avirulenta para neumonía y 671119 fue aislada de un bovino clínicamente sano en tanto HN06 (D) está asociada a Rinitis atrófica cuyo único factor suficiente para producir la enfermedad es PMT.

La recombinación homóloga está significativamente presente en *Pasteurella multocida*, afectando la reconstrucción filogenética. En el árbol genoma core mínimo se observa el clado D que agrupa principalmente cepas asociadas con cólera aviar, pero no en el árbol genoma core donde estas cepas muestran una topología distinta. El árbol genoma core mínimo muestra agrupamiento de cepas según su asociación a enfermedades específicas de una manera más clara que el árbol genoma core y MLST.



**Figura 8. árbol filogenómico de *P. multocida* del genoma core:** A) árbol filogenético del genoma core (1209 genes de copia única) y B) árbol filogenético de genoma core mínimo (648 genes sin evidencia de recombinación). Ambos árboles fueron contruidos usando máximo likelihood con 100 de Bootstrap en RAXML. Las letras A, B, C y D representan clados de cepas asociadas a cólera aviar (azul), enfermedades respiratorias (marrón), septicemia hemorrágica (morado) y cólera aviar (azul), respectivamente. En tanto cepas asociadas a humanos son más distantes (verde).

Pasteurellosis neumónica

Cólera aviar

Infecciones en humanos

Septicemia hemorrágica

Coriza

Avirulentas

Rinitis atrófica

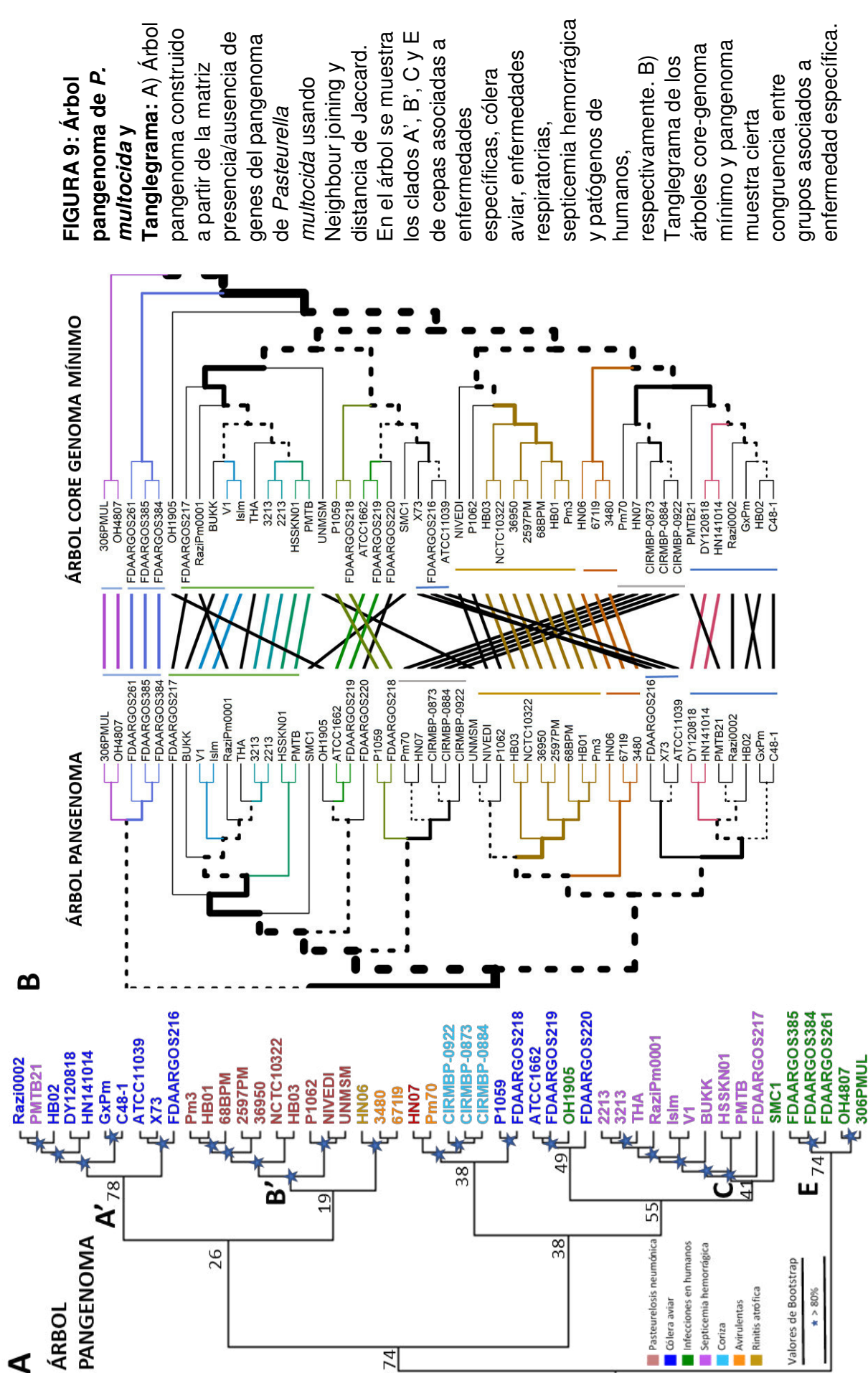
Valores de Bootstrap

★ > 80%

● > 70%

### 5.6.3. Reconstrucción filogenética del Pangenoma

Una de las principales fuerzas evolutivas del genoma bacteriano es la transferencia genética horizontal (TGH) que permite que individuos de una misma especie contengan un repertorio genético distinto. Este evento no puede ser representado en las filogenias moleculares a partir de secuencia, pero puede ser representado en un árbol filogenómico construido a partir del repertorio genético (presencia/ausencia de genes) (Kooning, 2016). Por ello se usó la matriz de presencia/ausencia del pangenoma para construir un árbol filogenético con el algoritmo *Neighbour joining* y la distancias Jaccard usando paquete *phangorn* y *ape* en el software R (Figura 9A). El árbol resultante genera un patrón sobre las bases del genoma accesorio y único y muestra una agrupación en clados de cepas asociadas a enfermedades específicas (en la figura 9A los colores de las cepas indican agrupación por enfermedad), esto significa que las cepas asociadas a una enfermedad específica conforman un clado debido a que comparten genes accesorios específicos. De manera resaltante se observa un nivel de congruencia entre el árbol pangenoma y el árbol genoma core mínimo a pesar de que ambos son conducidos por eventos distintos (TGH y mutaciones, respectivamente). Los clados A, B y C observados en el árbol genoma core mínimo también se observan en el árbol pangenoma pero con algunas diferencias que indican incluso un mayor agrupamiento de cepas asociadas. El clado A' incluye además las cepas ATCC11039, X73 y FADAARGOS219 también asociadas a cólera aviar. El clado B' incluye también las cepas NIVEDI y UNMSM asociadas a neumonías. De manera resaltante las cepas asociadas a ser patogénicas en humanos (FDAARGOS261, FDAARGOS384, FDAARGOS385, 306PMUL y OH4807) son las más distantes y se encuentran formando un clado monofilético (clado E). Para observar la congruencia entre el árbol genoma core mínimo y el árbol pangenómico se realizó un tanglegrama colocando ambos árboles frente a frente (Figura 9B) usando el paquete informático *dendextend* en R.



## 5.7. Diversidad del genoma accesorio

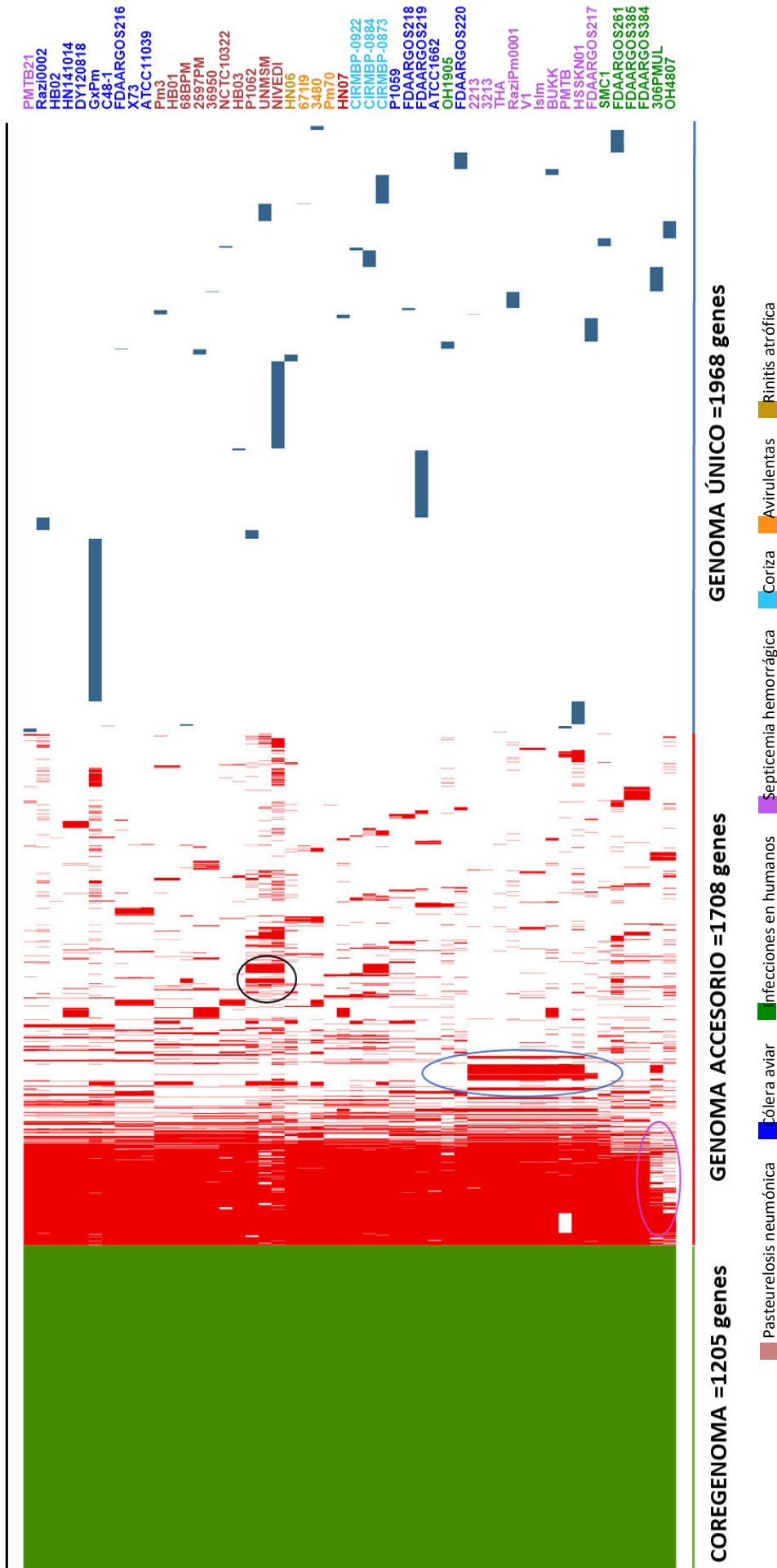
Según el análisis filogenético realizado existe una agrupación de cepas asociadas a enfermedades específicas, y esta disposición es concordante entre la filogenia del genoma core mínimo, y el pangenoma basado en presencia/ausencia de genes.

En la figura 10 se muestra un mapa de calor (*heatmap*) donde el genoma core (1205 genes), genoma accesorio (1708 genes) y genoma único (1968 genes) son representados por colores (verde, rojo y azul, respectivamente). Las filas son las cepas y las columnas son las familias génicas. La presencia del gen es representada por color (verde, rojo o azul según corresponda) y la ausencia de color blanco.

En este *heatmap* se observan algunos clúster de genes en el genoma accesorio que agrupa algunas cepas, por ejemplo, las cepas UNMSM, NIVEDI y P1062 asociadas a neumonías presentan regiones compartidas señaladas en círculo (morado), las cepas relacionadas a septicemia hemorrágica también poseen otra región compartida por todas estas cepas (círculo gris), mientras que las cepas relacionadas a ser zoonóticos (FDAARGOS 261, FDARRGOS 384, FDAARGOS 385, 306PMUL y OH4807) poseen regiones con genes ausentes (círculo negro).



## PANGENOMA = 4881 genes



**Figura 10: Heatmap del pangenoma de *Pasteurella multocida*.** Las filas representan los genomas y la presencia de genes (en columna) son mostrados por los colores verde, rojo y azul que representan el genoma core, genoma accesorio y genoma único, respectivamente. Los círculos muestran regiones que son comunes entre algunas cepas, el círculo negro muestra una región de genes compartido por cepas UNMSM, P1062 y NIVEDI (Todos de enfermedades respiratorias), el círculo azul encierra un conjunto de genes que son comunes a cepas de septicemia hemorrágica y el círculo morado que no presenta esos genes para los genomas asociados a enfermedades en humanos

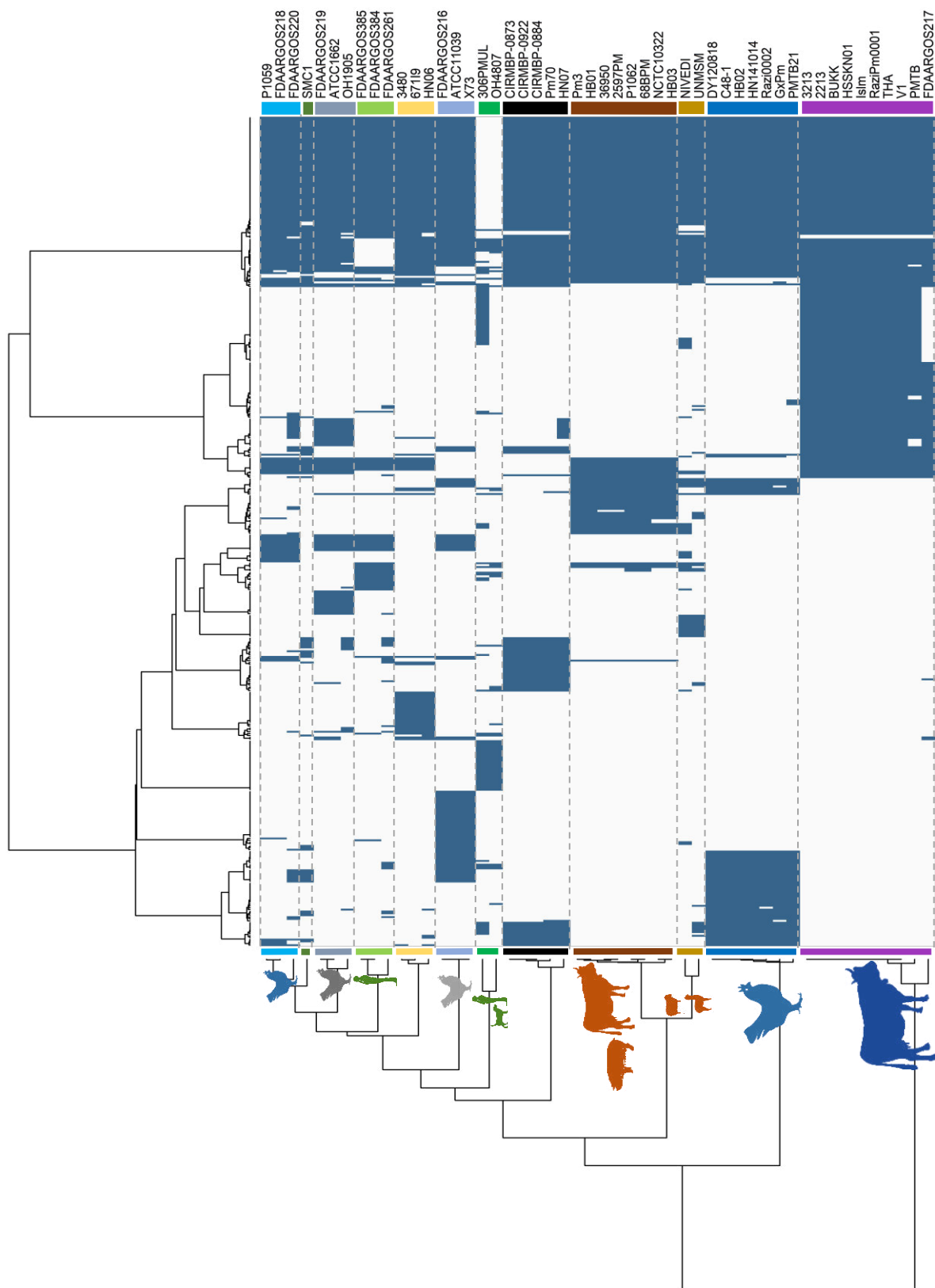


## 5.8. Identificación de genes relevantes del genoma accesorio que agrupan cepas relacionadas a enfermedad específica

Para identificar aquellos conjuntos de genes que son más importantes para la agrupación observada en el análisis filogenético del pangenoma y las señaladas en el *heatmap* de la figura 10, se utilizó un enfoque probabilístico bayesiano. Kpax2 es un paquete de R que agrupa datos de variables categóricas. En este caso consideramos a cada gen del genoma accesorio como una variable categórica (en total 1708 variables categóricas) con categorías binarias de (presencia/ausencia (1/0)). Como resultado se obtuvo 443 genes (26%) de los 1708 genes del genoma accesorio que son discriminatorios para los 12 grupos identificados por kpax2 (Tabla 3). A partir de estos genes se realizó un agrupamiento jerárquico usando *ggplot* en R con: *hclust(method=Ward.D); dist(method=Jaccard)*. La figura 11 muestra el *heatmap* construido con estos parámetros. Las agrupaciones obtenidas por kpax2 coinciden con las agrupaciones del árbol filogenético (cepas relacionadas a enfermedad específicas). Estos genes son descritos a detalles en la siguiente sección

**Tabla 3: Clústeres de cepas generado por kpax2 a partir de la matriz de presencia ausencia del genoma accesorio de *P. multocida*.**

clúster	Color	#genomas	Genomas	Enfermedad
1		10	3213, 2213, PMTB, RaziPm0001, BUKK, HSSKN01, THA, Islm, V1, FDAARGOS217	Septicemia H.
2		7	DY120818, C48-1, HB02, HN014014, Razi0002, GxPm, PMTB21	Cólera aviar
3		3	FDAARGOS2016, ATCC11039, X73	
4		3	FDAARGOS219, ATCC1662, OH1905	
5		3	P1059, FDAARGOS218, FDAARGOS220	
6		8	Pm3, HB01, 36950, 2597PM, P1062, 68BPM, NCTC10322, HB03	Pasteurellosis Neumónica
7		2	NIVEDI, UNMSM	
8		3	3480, 671119, HN06	Avirulentas y RA
9		5	CIRMBP-0873, CIRMBP-0922, CIRMBP-0884, Pm70, HN07	
10		2	306PMUL, OH4807	Relacionada a humanos
11		3	FDAARGOS261, FDAARGOS384, FDAARGOS385.	
12		1	SMC1	



**Figura 11: Heatmap de agrupamiento de cepas calculada por K-pax2 del genoma accesorio. Heatmap con 443 genes de mayor poder discriminatorio del genoma accesorio (1708 genes) llegan a formar 12 clústeres que coinciden con agrupamientos específicos a enfermedades. Los genomas que pertenecen a algún clúster son mostrados por una barra del mismo color según la tabla 3.**

## **5.9. Descripción funcional de genes presentes diferencialmente entre cepas asociadas a enfermedades específicas**

Los 443 genes identificados con k-pax2 como altamente discriminantes entre los 12 grupos fueron extraídos en sus secuencias de aminoácidos y re-anotados usando la base de datos nr para proteínas del NCBI. Luego son mostrados según su agrupamiento por enfermedad: septicemia hemorrágica, cólera aviar, pasteurelósisis neumónica y pasteurellosis en humanos (aislados de humanos y perros) en *heatmaps*.

### **5.9.1. Proteínas presentes de manera diferencial en cepas asociadas a septicemia hemorrágica**

Los genes que están presentes de manera diferencial en cepas asociadas a septicemia hemorrágica son mostrados en la figura 12 y en la tabla 4. Estas permiten el agrupamiento de las 10 cepas como fue observado en el árbol pangenoma y kpax2.

Setenta y seis proteínas fueron identificadas para el clúster de septicemia hemorrágica, la mayoría de los genes (50) han sido anotados como proteínas hipotéticas, 08 son proteínas de origen fágico, como endonucleasas y reguladoras de vías de reparación y replicación. Es conocido que los profagos integrados en los genomas de sus hospederos pueden influenciar su estilo de vida expresando reguladores globales importantes para la adaptación a nichos específicos (Braga *et al.*, 2018). Otras proteínas con función asociada a replicación, recombinación y reparación al daño de ADN como: proteína de reparación de alquilación de A+G, proteína de familia de nucleasas PD-(D/E)XK, helicasa dependiente de ATP y un factor de transcripción regulador de transcripción de familia XRE, también están presentes.

La presencia de proteínas relacionadas a metabolismo de lipopolisacáridos (LPS) como glicosiltransferasas y heptiltransferasas han sido identificadas presentes exclusivamente en estas 10 cepas. Trabajos previos reportan que proteínas LPS son importantes para la virulencia de *Pasteurella multocida* en especial en aquellas asociadas a septicemia hemorrágica (Harper *et al.*, 2004).

UDP-N-acetilglucosamina 2-epimerasa (no-hidrolizante) (BcbA), UDP-N-acetil-D-manosamina dehidrogenasa (BcbB), proteína de biosíntesis polisacárido capsular BcbE y proteína de biosíntesis polisacárido capsular BcbD son proteínas de metabolismo de glucosa que forman parte de la estructura capsular B de estas cepas y han sido reconocidos como importantes en la supervivencia en macrófagos (Boyce & Adler, 2000). Algunas proteínas han sido identificadas para ser virulentas en otros patógenos como la desoxiribonucleasa de la familia TatD que es importante en *Plasmodium* para contrarrestar las trampas extracelulares de neutrófilos (NETs) contrarrestando la eliminación inmune en estado asexual sanguíneo del patógeno (Chang *et al.*, 2016). Por otro lado, la proteína SppA, una proteasa bacteriana de péptido señal tiene la función de degradar péptidos antimicrobianos y contribuye a la resistencia a nisina en *Bacillus subtilis* (Kingston *et al.*, 2013). La proteína TolA, una proteína de membrana periplasmática de Gram-negativas es importante en la adherencia celular en muchas bacterias como *Salmonella* y *E. coli* ya que está involucrado en la producción de LPS (Morgan *et al.*, 2014), la proteína Fic de filamentación celular en *Vibrio parahaemolyticus* es responsable del colapso del citoesqueleto de actina de la célula hospedadora; al ser una proteína de respuesta a *stress*, se cree que la familia de proteínas Fic está involucrada en la regulación de vías de señalización de la célula del hospedero, como la cascada de señalización inmune mediando de esta forma la patogénesis bacteriana (Roy y Cherfils, 2015). Finalmente, la proteína CzcD es una proteína de membrana involucrada en el flujo de cationes metálicos como  $Zn^{+2}$  importante en la virulencia de *P. aeruginosa* y en la resistencia a antibióticos (Salusso y Raimunda, 2017).

**Tabla 4: Detalle de las proteínas presentes en cepas de *P. multocida* asociadas a septicemia hemorrágica.**

	Proteína	Función en virulencia
1	acetilgalactosaminil-proteoglican 3-beta-glucuronosiltransferasa	<b>Ausentes</b> Responsables de la formación del lipopolisacárido y tipo capsular A
2	Beta-1,4-galactosiltransferasa	
3	Proteína de la familia de lipopolisacarido heptosiltransferasa	
4	Desoxiribonucleasa de familia TatD	Implicado en la virulencia en <i>Plasmodium</i> , expresado en estado sanguíneo para evitar el ataque de neutrófilos (Chang <i>et al.</i> , 2016)
5	Proteína de la familia glicosiltransferasa grupo 2	Metabolismo de lipopolisacárido (Harper & Boyce 2017)
6	Proteína TolA de integridad de envoltura celular	Implicado en la virulencia de muchas bacterias Gram negativas. Como <i>Salmonella</i> Typhimurium y <i>E. coli</i> para adherencia celular (Morgan <i>et al.</i> , 2014)
7	UDP-N-acetilglucosamina 2-epimerasa (no-hidrolizante) BcbA	Formación del tipo capsular B (Townsend <i>et al.</i> , 2001) Esencial para la patogénesis de septicemia hemorrágica (Boyce & Adler, 2000)
8	UDP-N-acetil-D-manosamina deshidrogenasa BcbB	
9	proteína de biosíntesis polisacárido capsular BcbE	
10	proteína de biosíntesis polisacárido capsular BcbD	
11	Glicosil transferasa	Metabolismo de lipopolisacárido
12	Proteína SppA	Proteasa tipo IV inhibición del péptido antimicrobiano nisina segregado por macrófagos (Kingston <i>et al.</i> , 2013)
13	Proteína CzcD	Transportador de iones metálicos, especificidad por Zn <sup>2+</sup> , ayuda en la tolerancia a antibiótico en <i>P. aeruginosa</i> (Salusso y Raimuna, 2017)
14	Helicasa dependiente de ATP	Replicación, recombinación o reparación de ADN
15	Proteína de la familia de nucleasas PD-(D/E)XK	Replicación, recombinación o reparación de ADN
16	proteína Fic	Respuesta bacteriana al stress (Roy y Cherfils, 2015)
17	Proteína activadora de hemolisina	Hemolisina
18	HK97 family phage prohead protease phage portal protein phage head-tail adaptor phage gp6-like head-tail connector protein HNH endonuclease phage terminase small subunit P27 family AlpA family phage regulatory protein Transposase	Fago
19	Proteína hipotética (50)	Función desconocida

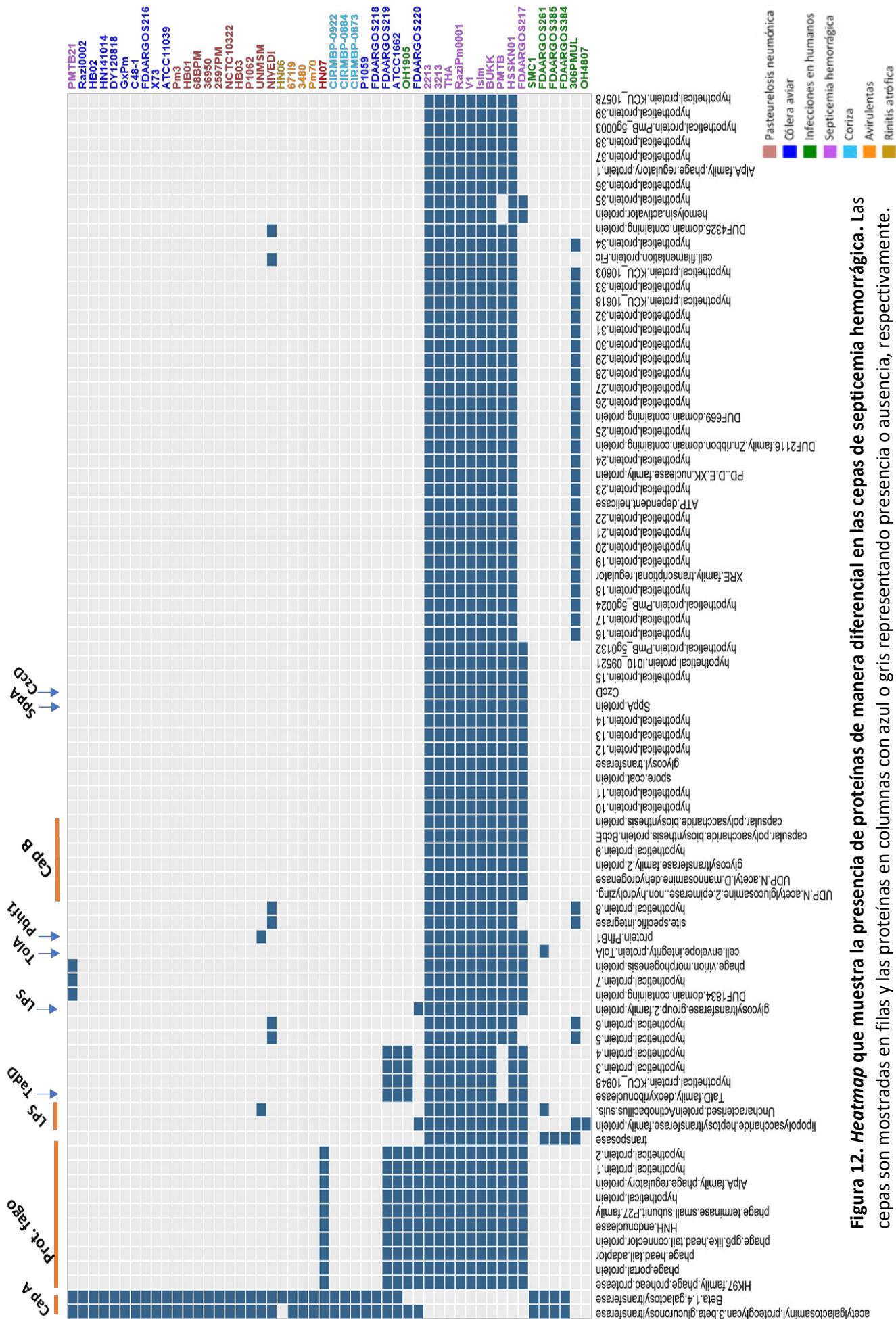


Figura 12. Heatmap que muestra la presencia de proteínas de manera diferencial en las cepas de septicemia hemorrágica. Las cepas son mostradas en filas y las proteínas en columnas con azul o gris representando presencia o ausencia, respectivamente.

### 5.9.2. Proteínas presentes de manera diferencial en cepas asociadas a pasteurellosis neumónica.

El clado identificado como B en el análisis filogenómico está conformado por las mismas cepas que los clústeres 6 y 7 agrupados por *kpax2* y son cepas asociadas a procesos neumónicos de distintos hospederos. En la figura 13 se muestra un *heatmap* con los genes presentes de manera diferencial en estas cepas identificadas por *kpax2*, en este mapa se puede diferenciar conjuntos de genes comunes a los 10 genomas, otro conjunto presente en 8 genomas y otro solo en 3 genomas (P1062, NIVEDI y UNMSM).

Resulta interesante la presencia de proteínas relacionadas al transporte y metabolismo de carbohidratos, en especial de tres proteínas del metabolismo de trehalosa: El operón trehalosa (Regulador Transcripcional (TreR), subunidad de transporte de trehalosa PTS IIBC (TreP) y alfa-fosfotrehalasa (TreA)) que está presente en las 10 cepas asociadas a enfermedades respiratorias, aunque la cepa UNMSM carece de *treA*. Este operón de metabolismo de trehalosa puede permitir a estas cepas poder importar y metabolizar este carbohidrato. En un reciente estudio se ha demostrado que un linaje de *Clostridium difficile*, muestra mayor virulencia que otros linajes gracias a que posee la capacidad de metabolizar bajas concentraciones de trehalosa. Este linaje ha sido asociado a recientes brotes nosocomiales de este patógeno en EE.UU. y Europa (Collins *et al.*, 2018).

Otras proteínas que están en las 10 cepas del clado asociado a procesos neumónicos, están relacionadas al transporte y metabolismo de carbohidratos: permeasa de transportador ABC C4-dicarboxilato, proteína de unión a sustrato de transportador ABC C4-dicarboxilato, reductasa 3-oxoacil-ACP, L-idonato 5-dehidrogenasa y permeasa TRAP (Moraes y Reithmeier, 2012). Aunque estas proteínas también están presentes en algunas cepas asociadas a cólera aviar. Asimismo, otros genes relacionados con el metabolismo energético como la proteína de la familia de pirina involucrada en la regulación de catabolismo de piruvato, oxidorreductasa NAD(P)H dependiente, un transportador de potasio, una proteína de familia SIR2 que actúa como una deacetilasa dependiente de NAD<sup>+</sup> (Anand *et al.*, 2017). Estas proteínas presentes principalmente en 9 cepas

aisladas de bovinos y cerdos y ovino, pero no a la cepa de alpaca (UNMSM). Finalmente, dos proteínas con roles en el metabolismo de hierro: Receptor TonB y Receptor de Sistema de Adquisición Hemo (Bosch *et al.*, 2002; Prado *et al.*, 2005) también están presentes diferencialmente en cepas asociadas a procesos neumónicos.

De manera resaltante, muchos genes de origen bacteriófago son comunes entre los genomas P1062, NIVEDI y UNMSM, especialmente en estos dos últimos, los cuales hace que sean más cercanos y formen un clúster distinto (clúster 7) en kpax-2.

**Tabla 5: Detalle de las proteínas presentes en cepas de *P. multocida* asociadas a pasteurelosis neumónica.**

	Proteína	Función en virulencia
1	Receptor TonB-dependiente	Proteína de membrana asociada al factor de virulencia TonB implicada en el metabolismo de hierro en <i>P. multocida</i> (Bosch <i>et al.</i> , 2002)
2	Permeasa de transportador ABC C4-dicarboxilato	Transportador de membrana para carbohidratos (Moraes y Reithmeier, 2012)
3	Proteína de unión a sustrato de transportador ABC C4-dicarboxilato	Transportador de membrana para carbohidratos (Moraes y Reithmeier, 2012)
4	3-oxoacyl-ACP reductasa	Biosíntesis de ácidos grasos
5	L-idonato 5-dehidrogenasa	Metabolismo de idonato (Moraes y Reithmeier, 2012)
6	Permeasa de transportador TRAP	Transportador de membrana para carbohidratos (Moraes y Reithmeier, 2012)
7	<b>Operón trehalosa</b> Regulador Transcripcional (TreR), Subunidad de transporte de trehalosa (TreP) y Alfa-fosfotrehalasa (TreA) Regulador Transcripcional (TreR), subunidad de transporte de trehalosa PTS IIBC (TreP) y alfa-fosfotrehalasa (TreA))	Metabolismo de trehalosa. Collins et. al., 2018 demostró que una dieta de trehalosa potencia la virulencia de ciertos brotes epidémicos de <i>Clostridium difficile</i>
8	Oxidoreductasa dependiente de NAD(P)H	Proteína de membrana oxidoreductasa
9	Transportador de Potasio	transportador de iones potasio en membrana
10	Proteína de familia Pirina	No caracterizado
11	Receptor de sistema de adquisición Hemo	Metabolismo de hierro (Prado <i>et al.</i> , 2005)
12	Proteína de la familia SIR2	Deacetilasa en bacterias, ¿regulación epigenética?) (Anand <i>et al.</i> , 2017)
13	Proteína relacionada a permeasa de Purina-Citocina	Transportador de bases nitrogenadas





### **5.9.3. Proteínas presentes de manera diferencial en cepas asociadas a cólera aviar**

El análisis filogenómico basado en contenido genético revela varios clados asociados a cólera aviar. kPax2 identifica 4 agrupaciones de cepas asociadas a cólera aviar en los que son encontrados una distribución diferente de proteínas (figura 14). El clúster 2 agrupa GxPm, C48-1, HN0141014, DY120818, HB02, Razi0002, PMTB2.1(Cepa de tipo A asociada a septicemia hemorrágica); el clúster 3 agrupa FDAARGOS216, X73, ATCC11039; el clúster 4 agrupa P1059, FDAARGOS218, FDAARGOS220 y el clúster 5 agrupa FDAARGOS219, ATCC1662, OH1905 (Cepa de tipo A asociada a infección en humanos).

Los clústeres 2 y 3 comparten una mayor cantidad de proteínas (que concuerda con el árbol filogenómico basado en pangenoma), estas proteínas son principalmente relacionadas a metabolismo y transporte de carbohidratos, en particular poseen los operones de arabinosa y alosa, para los cuales contienen: regulador transcripcional de familia MurR/RpiR (alsR), subunidad transportador de D-alosa (AlsB), proteína de unión a ATP de transportador ABC de alosa (AlsA), citocromo c (AlsC), D-alulosa-6-fosfato 3-epimerasa (AlsE), alosa kinasa (AlsK), ribosa 5-fosfato isomerasa B (RpbB) como parte del operón alosa y una ATPasa, fosfodiesterasa, L-arabinosa isomerasa (AraA), permeasa transportador ABC, permeasa transportador ABC, proteína de unión a sustrato de autoinductor de transportador ABC y regulador transcripcional de operón arabinosa (AraC) como genes del operón arabinosa.

Los clústeres 3, 4 y 5 comparten otro conjunto de genes relacionados al metabolismo de fucosa (L-fuculosa-fosfato aldolasa (FucA), L-fucosa mutarotasa (FucU), L-fuculokinasa (FutK), L-fucosa isomerasa (FucI), aldehído deshidrogenasa, Regulador transcripcional de la familia LacI, permeasa de transportador ABC (FucP) , proteína de unión a ATP de transportador de azúcar ABC y regulador transcripcional DeoR/GlpR (FucR).

De manera individual para cada clúster, el clúster 2 posee un conjunto de proteínas únicas donde resaltan aquellas relacionadas a metabolismo de carbohidratos: ATPasa, transportadores de membrana, etc. En el clúster 3

resaltan 5 proteínas relacionadas al metabolismo de galactitol y tagatosa: regulador transcripcional DeoR/GlpR (GatR), galactitol-1-fosfato 5-dehidrogenasa (GatD), subunidad transportadora de galactitol PTS IIC (GatC), subunidad transportadora de galactitol PTS IIB (GatB), subunidad transportadora de galactitol PTS IIA (GatA), subunidad tagatosa-bisfosfato aldolasa (GatZ) y tagatosa 1,6-difosfato aldolasa (GatY). El clúster 4 posee 5 genes que conforman el operón citrato para el metabolismo de citrato: proteína portadora de citrato liasa acil (CitD), citrato (pro-3S)-liasa subunidad beta (CitE), [citrato (pro-3S)-liase] ligasa (CitC), citrato liasa subunidad alfa (CitF) y Trifosforibosil-defosfo-CoA sintasa (CitG). Finalmente, el clúster 5 posee un conjunto de proteínas de la familia CRISPR-Cas.

Esta descripción da cuenta de que hay una presencia importante de proteínas asociadas a metabolismo de carbohidratos como fucosa, arabinosa, alosa, galactitol y citrato, distribuidos de manera distinta entre los cuatro clústeres, la presencia exclusiva de estos genes en cepas asociadas a cólera aviar sugiere que la adquisición de nuevas capacidades metabólicas puede ser importante para *Pasteurella multocida* y la patogénesis en cólera aviar.

**Tabla 6: Detalle de las proteínas presentes en cepas de *P. multocida* asociadas a cólera aviar**

	<b>Proteína</b>	<b>Función en virulencia</b>
1	Regulador transcripcional de Familia FadR Oxidorreductasa de familia gfo/Idh/MocA Permeasa de transportador de azúcar ABC Permeasa de transportador de azúcar ABC UgpC Oxidorreductasa de familia gfo/Idh/MocA Permeasa de transportador de carbohidratos ABC Regulador transcripcional de Familia LysR	Proteínas con función de transporte de carbohidratos
2	<b>Operón D-alosa</b> Regulador transcripcional de Familia MurR/RpiR (AlsR) Subunidad de transportador de D-alosa (AlsB) Proteína de unión de ATP de transportador ABC de alosa (AlsA) Citocromo c6 (AlsC) D-alulosa-6-fosfato 3-epimerasa (AlsE) alosa kinasa (AlsK) ribosa 5-fosfato isomerasa B (RpbB)	<b>Operón D-alosa</b> D-alosa es una hexosa que es convertido a fructosa-6-fostato por enzimas que pertenecen a este operón. Asimismo, las enzimas que conforman el sistema de transporte de D alosa, también pueden importar D-ribosa (pentosa) con una eficiencia similar (Kim <i>et al.</i> , 1997).
3	<b>Operón L-fucosa</b> aldehído deshidrogenasa Regulador transcripcional de Familia LacI Permeasa de transportador ABC (FucP) Proteína de unión de ATP de transportador ABC L-fuculosa-fosfato aldolasa (FucA) L-fucosa mutarotasa (FucU) L-fuculokinasa (FucK) L-fucosa isomerasa (FucI) Regulador transcripcional DeoR/GlpR (FucR)	<b>Operón L-Fucosa</b> L-fucosa es una hexosa que forma parte de muchos glicanos presentes en la superficie de células eucariotas principalmente células epiteliales de distintos sitios de mucosa (Pacheco <i>et al.</i> , 2012). La elevada concentración de L-fucosa en el nicho intestinal es un factor clave para la colonización bacteriana en bacterias patógenas como <i>C. jejuni</i> (Stahl <i>et al.</i> , 2010) o el probiótico <i>Lactobacillus rhamnosu</i> (Becerra <i>et al.</i> , 2015)
4	<b>Operón L-Arabinosa</b> ATPasa fosfodiesterasa L-arabinosa isomerasa (AraA) Permeasa de transportador ABC Permeasa de transportador ABC Proteína de unión a sustrato de transportador ABC autoinductor 2 Regulador transcripcional de operón arabinosa (AraC)	<b>Operón L-Arabinosa</b> L-arabinosa en una pentosa constituyente de la pared celular de algunas plantas y es fuente de carbono y energía (Schleif, 2010). Se ha demostrado que L-arabinosa, independientemente de su catabolismo, puede actuar como un inhibidor de la invasión de en <i>Salmonella</i> entérica a células intestinales (Lopez-Garrido <i>et al.</i> , 2015)

	Proteína	Función en virulencia
5	<b>Operón Citrato</b> [citrato (pro-3S)-liasa] ligasa (CitC) Proteína portadora de citrato liasa acil (CitD) citrato (pro-3S)-liasa subunidad beta (CitE) citrato liasa subunidad alfa (CitF) Trifosforibosil-defosfo-CoA sintasa (CitG) anión permeasa	<b>Operón Citrato</b> Solo algunos microorganismos tienen la capacidad de metabolizar citrato (García <i>et al.</i> , 2008). Tres genes estructurales <i>citDEF</i> y accesorios <i>citC</i> y <i>citG</i> son encontrados por ejemplo en <i>Haemophilus influenzae</i> y producen oxalacetato a partir de citrato (Sobczak y Lolkema, 2005)
6	<b>Operón Galactitol</b> Regulador transcripcional DeoR/GlpR (GatR) galactitol-1-fosfato 5-deshidrogenase (GatD) Subunidad transportadora de galactitol PTS IIC (GatC) Subunidad transportadora de galactitol PTS IIB (GatB) Subunidad transportadora de galactitol PTS IIA (GatA) Subunidad tagatosa-bisfosfato aldolasa (GatZ) tagatosa 1,6-difosfato aldolasa (GatY) Transportador de ion Proteína de unión a ATP	<b>Operón Galactitol</b> Galactitol (ductitol) es un polialcohol fermentado por muchas enterobacterias, como algunas cepas de <i>E. coli</i> (Shakeri-Garakani <i>et al.</i> , 2004). El operón Glactitol compuesto de 7 genes está involucrado en la utilización de galactitol y D-tagatosa. En un reciente trabajo se sugiere que los genes <i>gat</i> son importantes para la colonización intestinal en animales de producción (Nolle <i>et al.</i> , 2017)
7	CRISPR-associated endonuclease Cas2 CRISPR-associated endonuclease Cas1 CRISPR-associated endonuclease Cas2 hypothetical protein TIGR02584 family CRISPR-associated protein DUF1887 domain-containing protein CRISPR-associated protein Csx16 CRISPR system precrRNA processing endoribonuclease RAMP protein Cas6 hypothetical protein CRISPR-associated protein Csm4 type III-A CRISPR-associated RAMP protein Csm3 type III-A CRISPR-associated protein Csm2	
8	colanic acid biosynthesis glycosyltransferase WcaL ADN metiltransferasa Proteína hipotética Proteína asociada a nucleoide ADN glicosilasa Proteína hipotética Proteína hipotética Proteína hipotética transportador, parcial proteína de integridad de envoltura celular TolA nucleótido azúcar deshidrogenasa nucleótido azúcar deshidrogenasa Proteína hipotética Proteína hipotética glicosil transferasa familia 28 Proteína hipotética X73_00317	



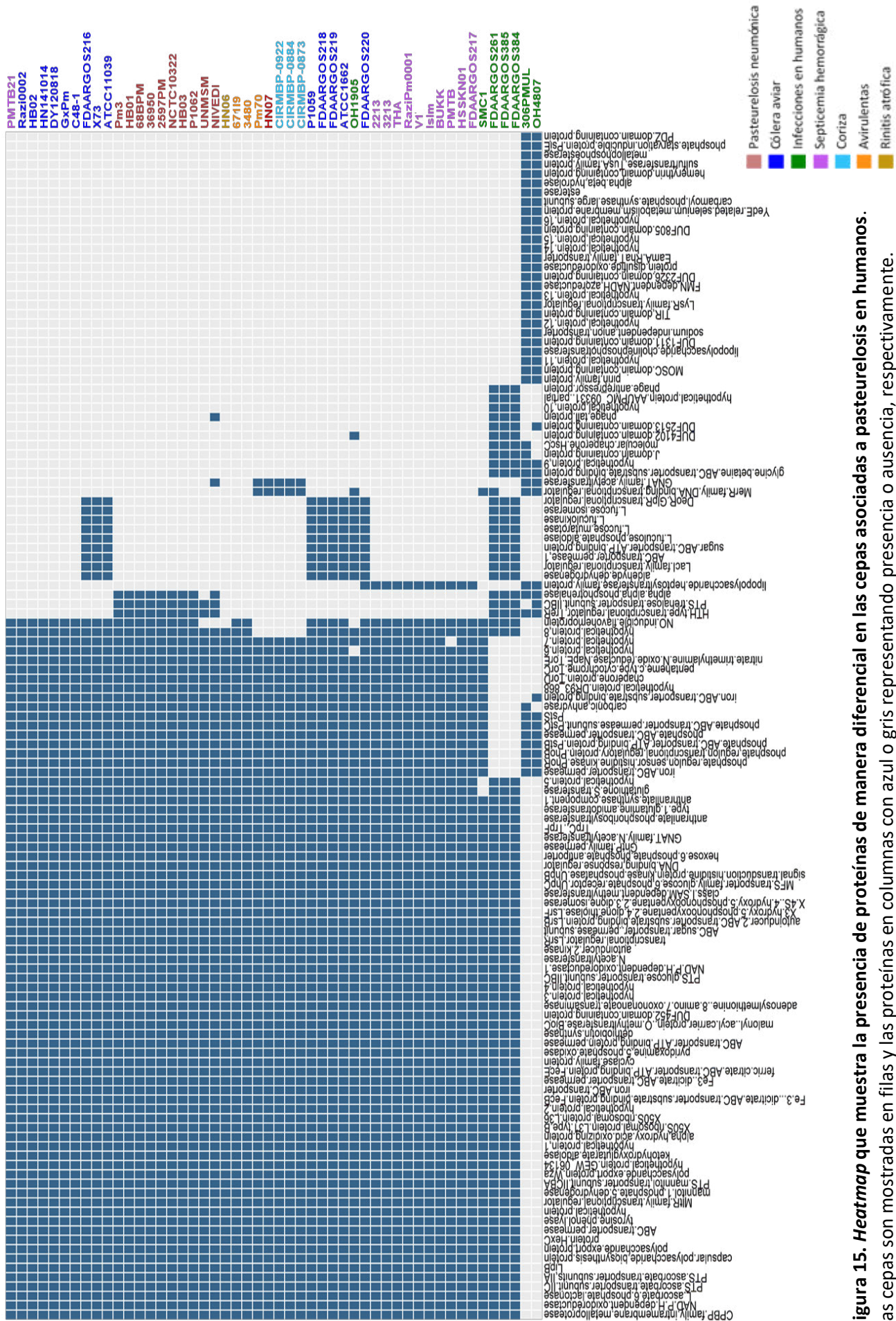


#### **5.9.4. Proteínas presentes/ausentes de manera diferencial en cepas asociadas a pasteurellosis en humanos**

En el análisis filogenómico, las cepas más divergentes son cepas relacionadas infecciones en humanos (306PMUL, FDAARGOS261, FDAARGOS384, FDAARGOS385) y perro (OH4807 – con potencial zoonótico) otras dos cepas aisladas de humano (SMC1) y perro (OH1907) se encuentran en otros clados. Lo más característico de estas 5 cepas es la reducción en el contenido genético.

K-pax2 identifica dos clústeres con genes ausentes/presentes de forma diferente: (FDAARGOS261, FDAARGOS384 y FDAARGOS385) y (OH4807 y 306PMUL). Los cuales carecen de 8 y 58 proteínas que se encuentran presentes en el resto de cepas analizadas. En común ambos muestran la ausencia de 6 genes de las cuales 4 son parte del operón TorECA y 2 son proteínas hipotéticas (Figura 15).







## 6. DISCUSIÓN

### 6.1. El análisis pangenómico muestra una amplia variación en el contenido genético de *P. multocida*

Se ha denominado a *P. multocida* como un patógeno enigmático debido a que posee la capacidad de infectar a un amplio rango de hospederos animales incluido el humano y producir en ellos enfermedades muy distintas, a pesar de su impacto en la economía y salud se conoce poco sobre las bases genéticas que subyacen esta capacidad (Wilkie *et al.*, 2012, Wilson y Ho, 2013). En el presente trabajo se elaboró el pangenoma de *Pasteurella multocida* a partir de 50 genomas, el mayor número de genomas analizados hasta el momento en esta especie, y se resalta su diversidad patogénica, metabólica y genética. Este enfoque sumado al análisis filogenómico da luces sobre la diversidad y evolución del genoma de esta bacteria que pueden explicar su adaptación y especialización al amplio repertorio de hospederos y enfermedades al que está asociado.

La identificación y caracterización de la totalidad de genes del pangenoma de una especie es una poderosa herramienta para estudiar la diversidad genómica (Tetellin *et al.*, 2005). Con el análisis de 50 genomas, *P. multocida* posee un pangenoma de 4 881 genes del cual solo 1 205 genes son compartidos por todas las cepas. El tamaño del pangenoma es muy superior al reportado (3585 genes) por Hurtado *et al.* (2018) donde solo 23 genomas fueron usados. El genoma core, como en casi todas las especies bacterianas (McInerney *et al.*, 2017), está principalmente representado por genes con funciones esenciales (77% para *P. multocida*) como metabolismo energético y regulación de la expresión. En promedio, el genoma core representa el 55% de un genoma promedio de *P. multocida* (2200 genes). Esta porción del genoma core es comparativamente alta considerando algunos organismos Gram-negativos como *Clostridium perfringens* (el genoma core representa 25 %) (Kiu *et al.*, 2017) o *E. coli* (el genoma core representa el 40%) (Rasko *et al.*, 2008), pero similar a *S. aureus* (el genoma core representa el 56%) (Bosi *et al.*, 2016).

La expansión del pangenoma de *P. multocida* es catalogado como “abierto”, lo que significa que teóricamente el repertorio genético de esta especie no tiene límite. De manera que con 50 genomas analizados el genoma accesorio y único representa el 75% del tamaño del pangenoma, y 45% de un genoma promedio de *P. multocida*. La anotación funcional tanto del genoma accesorio y único es enriquecido principalmente por proteínas de función desconocida y secuencias profago. El comportamiento abierto del pangenoma y el alto contenido de genes hipotéticos y secuencias profago también es observado en otras especies incluyendo *E. coli* y *S. aureus* (Price *et al.*, 2008; Bosi *et al.*, 2016), esto indica que la transferencia genética horizontal en *P. multocida* tiene un importante efecto en la dinámica evolutiva del genoma de esta especie como ocurre con las otras especies (Poltz *et al.*, 2013). Existe variada evidencia que sugieren que la transferencia genética horizontal en *Pasteurella multocida* es importante en su virulencia, a partir del análisis de genomas individuales muchos trabajos han descrito la presencia de islas de patogenicidad, resistencia y secuencias profago: la toxina PMT que produce rinitis atrófica es codificada por un bacteriófago lisogénico (Pullinger *et al.*, 2004), el elemento conjugativo integrativo (ICEPmu1) codifica para genes de resistencia a antibióticos en la cepa 36950 (Brenner *et al.*, 2012), o la cepa UNMSM aislada de alpaca posee 6 profagos predichos *in-silico* (Hurtado *et al.*, 2017).

El tipo de pangenoma (abierto o cerrado) se suele relacionar al estilo de vida de la bacteria, es así que un pangenoma abierto es característico de especies que poseen la habilidad de colonizar diferentes nichos. Esto es evidente si consideramos que la expansión del pangenoma significa tener una gran capacidad para adquirir nuevos genes, muchos de los cuales pueden portar elementos adaptativos. Puesto que *P. multocida* posee una alta versatilidad al ser capaz de infectar un amplio rango de hospederos es posible que esta capacidad subyace en la variabilidad genómica observada.

## **6.2. El análisis filogenético revela que cepas asociadas a enfermedades específicas conforman clados separados.**

El análisis de MLST ha sido ampliamente usado para estudiar la diversidad genética y la estructura poblacional de patógenos (Jolley y Maiden, 2014). De

los 50 genomas analizados por MLST *in-silico* se identificaron un total de 15 ST distintos, 10 genomas asociados a septicemia hemorrágica pertenecen al ST-122 y 8 genomas asociados a neumonías pertenecen al Complejo Clonal CC-ST-13. El comportamiento clonal de *P. multocida* ha sido motivo de estudio por Hotchkiss *et al.* (2013) quienes mediante análisis de MLST mostraron que cepas asociadas a enfermedades respiratorias en bovinos mantenía un comportamiento clonal incluyendo aislados de distintos tiempos y ubicaciones geográficas. En tanto que la presencia de cepas de SH es endémica del norte de África (serotipo E) y Asia (Serotipo B), donde poca diversidad de ST es observada y ST122 es el genotipo más encontrado (Petersen *et al.*, 2014; Sarangi *et al.*, 2014). Actualmente, gracias a recientes estudios genómicos, se entiende que la clonalidad, definida como la relación entre la tasa de recombinación y mutación ( $r/m$ ) o el balance entre la herencia vertical y horizontal, es una característica dinámica que puede variar a través del tiempo y entre poblaciones, como es el caso de ciertas bacterias como *Vibrio parahaemolyticus* o *Yersinia pestis* los cuales en una etapa patogénica pueden experimentar una explosión clonal (Bobay *et al.*, 2015; Shapiro, 2015). En un reciente trabajo Rimac *et al.* (2017) describen mediante BOX-PCR y ERIC-PCR la diversidad de *Pasteurella multocida* aisladas de un brote de neumonía en alpacas, en donde la diversidad de los aislados es casi nula, y en el que se puede sugerir que en una etapa patogénica *P. multocida* experimenta una expansión clonal como también se ha sugerido para un brote de cólera aviar (Petersen *et al.*, 2001). Sin embargo, el análisis de MLST solo considera 7 genes, es decir solo el 0.02 % del total de genes de un genoma promedio de *P. multocida*.

La recombinación es el mecanismo usado para incorporar nuevo material genético adquirido por TGH en el genoma, pero también puede introducir variación en genes *housekeeping* mediante recombinación homóloga (Didelot y Maiden 2010; Hanage, 2016). A nivel genómico se observa que la influencia de la recombinación en el genoma core de *P. multocida* es significativo, casi la mitad de genes core (519 de 1205 genes) han sufrido recombinación homóloga produciendo inconsistencias filogenéticas cuando es comparado con la filogenia de genes sin evidencia de recombinación (genoma core mínimo). El porcentaje de genes afectados por este fenómeno depende de la especie bacteriana, siendo

más frecuente por ejemplo en *Helicobacter pylori* (una especie denominada como panmíctica) y casi raro en *Mycobacterium tuberculosis* (Vos y Didelot, 2009; Hanage *et al.*, 2016). Aunque el 43% de genes del genoma core muestra evidencia de recombinación en al menos una cepa, esta no afecta las relaciones evolutivas entre los tres clados asociados a enfermedades específicas (clados A, B y C). Esto puede indicar que la mayoría de eventos de recombinación ocurre dentro de los clados y no entre cepas de distintos clados. Esto coincide con el concepto de que la recombinación proporciona una fuerza cohesiva que mantiene estos clados como unidades de diversidad (Shapiro, 2015).

La reconstrucción filogenómica a partir del genoma core es generalmente usada para inferir las relaciones evolutivas y es la primera herramienta para evaluar la diversidad (Treangen *et al.*, 2014; Millan-Aguiñaga *et al.*, 2017). El árbol construido a partir del genoma core mínimo de *P. multocida* teóricamente es la representación más fiel de la herencia vertical dentro de esta especie donde la interferencia de eventos de recombinación ha sido eliminada. Este árbol (figura 8B) muestra 4 clados (A, B, C y D) que presentan una clara asociación a enfermedades, una evidencia de que existe una especialización por parte de cepas de *P. multocida* para producir enfermedades específicas. Las variaciones adaptativas han sido mostradas usando análisis filogenómico en una variedad de especies bacterianas. Por ejemplo, Yue *et al.*, 2015 demostró cómo SNPs no sinónimos patoadaptativos en el genoma core de *Salmonella enterica* serovar Typhimurium están asociados a genes relacionados con la colonización como adhesinas así como otras distintas señales de SNPs asociados a hospedero específico. Por otro lado, la evolución a formas más virulentas también ha sido demostrada mediante filogenia del genoma core en *Leptospira spp.* (Thibeaux *et al.*, 2018).

Recientes estudios han mostrado que la concordancia entre la filogenia basada en SNPs del genoma core y regiones del pangenoma de presencia/ausencia es alta. Por ejemplo, McNally *et al.* (2016) estudiaron el linaje ST131 de *E. coli* donde los genomas core y accesorio muestran una alta concordancia y el análisis combinado de ambos permite analizar la evolución de *E. coli* en una resolución que no es posible si el análisis se restringe a una porción del genoma.

En un reciente trabajo Laing *et al.* (2017) elaboraron el pangenoma de *Salmonella entérica* analizando 4893 genomas donde encontraron una alta concordancia entre la filogenia basada en SNPs y una filogenia basada en presencia/ausencia de todos los genes del pangenoma. En el caso de *P. multocida*, como demuestra este trabajo, muestra cierta concordancia entre su genoma core y accesorio. El análisis del core y pan-genoma de *P. multocida* coinciden en un agrupamiento de cepas según enfermedad específica mostrado en los clados A, B y C de los árboles filogenéticos. Esto sugiere dos cosas, primero que la adquisición de genes mediante TGH no es azarosa, sino que la selección dentro de nichos específicos establece un complemento de genes que permite la especialización de cepas de *P. multocida* para ese ambiente específico. Segundo, esto es congruente con el concepto de que una bacteria debe ser “genómicamente coherente”, y que esta unidad puede estar conducida por tasas “razonables” de recombinación (Martinnen *et al.*, 2015).

### **6.3. El genoma accesorio presenta genes distribuidos diferencialmente en cepas de *P. multocida* asociadas a distintas enfermedades.**

Algunos trabajos han demostrado que la información para la adaptación a un nicho específico está almacenada en el genoma accesorio (Carter, 2016), por ejemplo, *Campylobacter fetus* posee linajes que infectan humanos y bovinos que contienen distintos patrones de genes del genoma accesorio que contribuyen a su especificidad por hospedero y puede ser explicado por evolución adaptativa (Iraola *et al.*, 2017). En esa misma línea, la filogenia del pangenoma de *P. multocida* muestra una segregación relacionada con la especialización a enfermedad específica lo que sugiere una presencia diferencial de genes que determina esta distribución según enfermedad. Otro enfoque de agrupamiento de tipo bayesiano revela 12 clústeres que coincide con el patrón filogenético del pangenoma soportado por 443 genes presentes diferencialmente en clústeres asociados a enfermedades específicas (figura 11). El análisis filogenético y la identificación de conjuntos de genes presentes exclusivamente en cepas asociadas a enfermedades específicas permite suponer que estos genes pueden ser responsables de la variada patogénesis de *P. multocida*. Además, la anotación de cada uno de estos genes revela diferencias en cuanto a la función

que cumplen los grupos de genes presentes de manera diferencial en estos clústeres de cepas asociadas a enfermedades específicas.

Las cepas asociadas a septicemia hemorrágica poseen 77 proteínas casi exclusivamente, entre ellas resaltan proteínas responsables de la síntesis de lipopolisacáridos y cápsula B (7 genes). Ambos importantes para invasión y resistencia a macrófagos (Harper *et al.*, 2004; Annas *et al.*, 2015). Previos trabajos demostraron que los LPS de *P. multocida* tipo B:2 desestabiliza los filamentos de actina en células endoteliales de aorta bovina (CEAB), que puede permitirles translocarse del tracto respiratorio al torrente sanguíneo, una propiedad que es más importante en serotipo B:2 que en el serotipo A:3 (Othman *et al.*, 2012; Yap *et al.*, 2018). Las otras proteínas presentes en este grupo de organismos como TolA y FicC están relacionadas a adherencia celular y modificación de filamentos de actina en la célula hospedero, en tanto TatD y SppA son proteínas de defensa frente al ataque inmune. Aunque la mayoría de proteínas son hipotéticas, las pocas descritas funcionalmente están involucradas en adherencia, invasión y defensa al sistema inmune.

En cambio, interesantemente, las cepas asociadas a cólera aviar, presentan otras proteínas como operones asociados a metabolismo de carbohidratos (D-alosa, galactitol (ductitol), tagatosa, L-fucosa, L-arabinosa y citrato) casi todas exclusivas de estas cepas. En tanto las cepas asociadas a pasteurelosis neumónica de cerdos, bovinos, ovinos y alpaca poseen diferencialmente solo un operón de trehalosa y otros genes relacionados con metabolismo energético. *P. multocida* es patógeno primario en cólera aviar de aves y patógeno secundario en pasteurelosis neumónica; pero en ambos casos está involucrado el mismo serogrupo A (Wilson *et al.*, 2012). El análisis filogenético demuestra que ambas enfermedades están asociadas a cepas tipo A filogenéticamente distintas y con presencia diferencial de genes. Aunque, Johnson *et al.* (2013) encontraron genes para metabolismo de L-fucosa y D-alosa exclusivamente en cepas virulentas de aves X73, P1059 de *P. multocida*, la presencia de genes relacionados a metabolismo de múltiples fuentes de carbono no ha sido observada de manera amplia en cepas de cólera aviar sino con este trabajo. Sellyei *et al.* (2015) revela que cepas de *P. multocida* aisladas de pulmón de

bovinos con procesos neumónicos presentaban metabolismo de trehalosa de manera diferencial a cepas aisladas de cavidad nasal, en coincidencia, la presencia de un operón de trehalosa se generaliza a cepas de otros hospederos de la misma enfermedad. La importancia que puede tener la capacidad de metabolizar trehalosa para la virulencia ha sido recientemente demostrado en *Clostridium difficile*, en el que un linaje que posee la capacidad de metabolizar bajas concentraciones de trehalosa gracias a un clúster de genes adquiridos fue responsable de brotes epidémicos nosocomiales con un incremento en la severidad de la enfermedad (Collins *et al.*, 2018). Estas evidencias sugieren la idea de que los genes metabólicos pueden estar desempeñando un papel importante en la virulencia en *P. multocida*.

#### **6.4. La presencia de proteínas con función de transporte y metabolismo de carbohidratos pueden conferir capacidad para colonizar nuevos hospederos**

Se reconoce que el genoma core de bacterias está principalmente compuesto de genes metabólicos, sin embargo, hay un creciente número de estudios que demuestran que genes metabólicos y de transporte son también parte del genoma accesorio y que estos pueden contribuir significativamente a la diversidad observada en las poblaciones bacterianas (Watkins *et al.*, 2016). Consecuentemente, el genoma accesorio de *P. multocida* muestra un porcentaje relativamente mayor de genes involucrados en el metabolismo de carbohidratos que el genoma core, según el análisis de anotación funcional por COG.

Son conocidos los clásicos factores de virulencia identificados en *P. multocida* que incluyen cápsula, lipopolisacáridos, proteínas de adhesión a superficie, de adquisición y regulación de hierro (Harper *et al.*, 2006). Sin embargo, las bacterias patógenas se encuentran con una variedad de ambientes durante su proceso de infección a los cuales debe adaptarse haciendo uso de fuentes alternativas de nutrientes, fenómeno denominado como ‘virulencia nutricional’ (Kwaik y Bumann, 2013). Esta hipótesis puede aplicarse al caso de *P. multocida*, que produce diferentes enfermedades enfrentándose a distintos ambientes en cada uno de sus hospederos: ya sea a nivel intracelular o extracelular, en mucosas, sangre, tracto respiratorio, tracto digestivo o en la piel.

La capacidad para obtener nutrientes es un campo de batalla principal de la interacción patógeno-hospedero, y es uno de los aspectos fundamentales en las enfermedades infecciosas donde los patógenos enfrentan complejas dinámicas nutricionales en microambientes del hospedero (Rohmer *et al.*, 2011). La expansión de las capacidades metabólicas puede ser una estrategia de adaptación a nicho considerando que fuera de las células la concentración de glucosa es baja con muchos otros carbohidratos ampliamente abundantes (Passalacqua *et al.*, 2016). En cólera aviar el ingreso primario de *P. multocida* se da a través de las vías respiratorias hacia el pulmón. Sin embargo, la supervivencia de diferentes cepas de *P. multocida* en intestino y subsecuente colonización de la mucosa del íleon y yeyuno origina la posibilidad de ingreso a través del tracto digestivo como una vía alternativa (Lee *et al.*, 2000; Mbuthia *et al.*, 2011). Estas cepas actúan a nivel extracelular en la mucosa respiratoria e intestinal, en cada cual los requerimientos energéticos son variables, enfrentándose incluso a la microbiota residente. Por tanto, la presencia adicional de genes y operones con función de transporte y metabolismo de carbohidratos en el genoma accesorio puede favorecer la colonización y establecimiento de cepas de *P. multocida* asociadas a cólera aviar en la superficie del tejido respiratorio o intestinal. Esta estrategia es observada en *Streptococcus mutans*, agente etiológico de las caries dentales, que se distingue de la microbiota oral residente no patógena por tener un rango metabólico grande y flexible con respecto a la adquisición de nutrientes. Como la dieta humana ha cambiado con el tiempo para volverse más rica en carbohidratos (especialmente sacarosa), también ha evolucionado la capacidad fisiológica de *S. mutans* (Moye *et al.*, 2014).

Las cepas de *P. multocida* relacionadas con Infecciones en humanos producto de la interacción con animales domésticos como perros (FDAARGOS261, FDAARGOS384, FDAARGOS385, OH4708 y 306PMUL) son más distantes filogenéticamente, además, comparado con los genomas de cepas de otros hospederos animales, resalta la carencia de un conjunto de genes. El operón *TorECA* involucrado en la respiración anaeróbica está ausente en las 5 cepas asociadas a infecciones en humanos (Ansaldi *et al.*, 2007). Un conjunto de 54 genes está ausente en cepas OH4708 y 306PMUL y cumplen funciones de



metabolismo energético, entre ellos operón ascorbato, manitol, etc. Existen reportes de *P. multocida* atípicos aislados de infecciones en humanos producto de mordeduras por perros y gatos, que muestran incapacidad de metabolizar sucrosa o manitol (Oberhofer, 1981; Christensen *et al.*, 2005).

Por tanto, el análisis pangenómico realizado en 50 cepas de *P. multocida* demuestra una agrupación de cepas por enfermedad específica en casos de septicemia hemorrágica, cólera aviar, pasteurelisis neumónica y relacionadas a infecciones en humanos tanto a nivel del genoma core y genoma accesorio. Dentro de estas últimas resalta la presencia de genes de metabolismo y transporte de carbohidratos que pueden estar relacionados con la patogenicidad y la capacidad de colonizar distintos ambientes en cada hospedero. Este análisis también demuestra que *P. multocida* presenta una gran diversidad genómica con fuerte influencia de la transferencia genética horizontal y recombinación en la evolución de esta especie.

## 7. CONCLUSIONES

Se logró caracterizar el pangenoma de *Pasteurella multocida* el cual muestra una importante variación en el contenido genético y donde el genoma accesorio y genoma único representan juntos el 75% con elevada presencia de proteínas hipotéticas y elementos móviles como resultado de transferencia genética horizontal.

Las cepas de *P. multocida* asociadas a enfermedades específicas como septicemia hemorrágica, cólera aviar, pasteurelosis neumónica e infecciones en humanos conforman clados separados en el análisis filogenómico y presentan genes accesorios de manera diferencial.

Determinados genes accesorios de transporte y metabolismo de carbohidratos (operones para metabolismo de alosa, fucosa, trehalosa, arabinosa, galactitol, tagatosa y citrato) están distribuidos de manera diferencial entre las cepas de *P. multocida* y pueden estar involucrados en la variada patogenicidad de esta bacteria al conferir nuevas capacidades para colonizar distintos ambientes en sus hospederos.

## 8. RECOMENDACIONES

Este trabajo puramente bioinformático predice genes que pueden estar involucrados en la patogénesis de las enfermedades que causa *P. multocida*, esto debe ser contrastado experimentalmente, realizando estudios de expresión *in vitro* o *in vivo* y evaluando las capacidades metabólicas de cada cepa.

Es necesario utilizar una mayor cantidad de genomas considerando la representatividad de cepas a nivel geográfico y de hospedero para corroborar la hipótesis de asociación de cepas especializadas a enfermedades específicas.

Para precisar la asociación contenido genético-enfermedad se pueden utilizar otros enfoques como GWAS (*Genome- wide association study*), que es ampliamente usado en genética humana.

## 9. REFERENCIA BIBLIOGRÁFICAS

- Abrahante, J. E. *et al.* Draft Genome Sequences of Two Virulent Serotypes of Avian *Pasteurella multocida*. *Genome Announc.* **1**, 1–2 (2013).
- Altschup, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic Local Alignment Search Tool. *J. Mol. Biol.* **215**, 403–410 (1990).
- Anand, C., Garg, R., Ghosh, S. & Nagaraja, V. A Sir2 family protein Rv1151c deacetylates HU to alter its DNA binding mode in *Mycobacterium tuberculosis*. *Biochem. Biophys. Res. Commun.* **493**, 1204–1209 (2017).
- Annas, S., Jesse, F. F. A. & Zunita, Z. Comparative clinicopathological changes in buffalo and cattle following infection by *Pasteurella multocida* B:2. *Microb. Pathog.* **88**, 94–102 (2015).
- Ansaldi, M., Théraulaz, L., Baraquet, C., Panis, G. & Méjean, V. Aerobic TMAO respiration in *Escherichia coli*. *Mol. Microbiol.* **66**, 484–494 (2007).
- Becerra, J. E., Yebra, M. J. & Monedero, V. An L -Fucose Operon in the Probiotic *Lactobacillus rhamnosus* GG Is Involved in Adaptation to Gastrointestinal Conditions. *Appl. Environ. Microbiol.* **81**, 3880–3888 (2015).
- Binnewies, T. T. *et al.* Ten years of bacterial genome sequencing: comparative-genomics-based discoveries. *Funct. Integr. Genomics* **6**, 165–185 (2006).
- Blackall, P.J. & Miflin, J. K. Identification and typing of *Pasteurella multocida*: A review. *Avian Dis.* **29**, 271–287 (2000).
- Bobay, L.-M., Traverse, C. C. & Ochman, H. Impermanence of bacterial clones. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 8893–900 (2015).
- Bosch, M. *et al.* *Pasteurella multocida* *exbB*, *exbD* and *tonB* genes are physically linked but independently transcribed. *FEMS Microbiol. Lett.* **210**, 201–208 (2002).
- Bosi, E. *et al.* Comparative genome-scale modelling of *Staphylococcus aureus* strains identifies strain-specific metabolic capabilities linked to pathogenicity. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E3801-9 (2016).
- Boyce, J. D., Seemann, T., Adler, B. & Harper, M. in *Pasteurella Multocida* 23–38 (2012). doi:10.1007/82
- Boyce, J. D., Harper, M., Wilkie, W. & Adler, B. in *Pathogenesis of Bacterial Infections in Animal* 325–346 (2010).
- Boyce, J. D. & Adler, B. The Capsule Is a Virulence Determinant in the Pathogenesis of *Pasteurella multocida* M1404 (B:2). *Infect. Immun.* **68**, 3463–3468 (2000).
- Braga, L. P. P. *et al.* Bacterial Diversification in the Light of the Interactions with Phages: The Genetic Symbionts and Their Role in Ecological Speciation. *Front. Ecol. Evol.* **6**, 1–12 (2018).
- Bruen, T. PhiPack: PHI test and other Test of Recombination. 1–8 (2005).

- Capella-Gutierrez, S., Silla-Martinez, J. M. & Gabaldon, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
- Carter, M. Q. Decoding the Ecological Function of Accessory Genome. *Trends Microbiol.* **25**, 1–2 (2017).
- Chakraborty, S., Kloos, B., Harre, U., Schett, G. & Kubatzky, K. *Pasteurella multocida* Toxin Triggers ranKl-independent Osteoclastogenesis. *Front. Immunol.* **8**, 1–16 (2017).
- Chang, Z. *et al.* The TatD-like DNase of *Plasmodium* is a virulence factor and a potential malaria vaccine candidate. *Nat. Commun.* **7**, 1–9 (2016).
- Christensen, H. & Bisgaard, M. in *Prokaryotes* 1062–1090 (2006).
- Christensen, H., Bisgaard, M., Angen, Ø., Frederiksen, W. & Olsen, J. E. Characterization of Sucrose-Negative *Pasteurella multocida* Variants, Including Isolates from Large-Cat Bite Wounds. *J. Clin. Microbiol.* **43**, 259–270 (2005).
- Christensen, J. P. & Bisgaard, M. Fowl cholera Description of the aetiological agent. *Rev. sci. tech. Off. Int. Epiz.* **19**, 626–637 (2000).
- Christenson, E. S., Ahmed, H. M. & Durand, C. M. Grand Round *Pasteurella multocida* infection in solid organ transplantation. *Lancet Infect. Dis.* **15**, 235–240 (2015).
- Clark, K. *et al.* GenBank. *Nucleic Acids Res.* **44**, D67–D72 (2016).
- Collins, J. *et al.* Dietary trehalose enhances virulence of epidemic *Clostridium difficile*. *Nature* **553**, 291–294 (2018).
- Dabo, S. M., Taylor, J. D. & Confer, A. W. *Pasteurella multocida* and bovine respiratory disease. *Anim. Heal. Res. Rev.* **8**, 129–150 (2008).
- Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* **27**, 1164–1165 (2011).
- Davies, R. L. Genetic diversity among *Pasteurella multocida* strains of avian, bovine, ovine and porcine origin from England and Wales by comparative sequence analysis of the 16S rRNA gene. *Microbiology* **150**, 4199–4210 (2004).
- Descamps, S., Jenouvrier, S., Gilchrist, H. G. & Forbes, M. R. Avian Cholera, a Threat to the Viability of an Arctic Seabird Colony? *PLoS One* **7**, (2012).
- Dey, S. *et al.* Comparative sequence analysis of 16S rRNA gene of *Pasteurella multocida* serogroup B isolates from different animal species. *Res. Vet. Sci.* **83**, 1–4 (2007).
- Didelot, X. & Maiden, M. C. J. Impact of recombination on bacterial evolution. *Trends Microbiol.* **18**, 315–322 (2010).
- Du, H. *et al.* Comparative Genomics Analysis of Two Different Virulent Bovine *Pasteurella multocida* Isolates. *Int. J. Genomics* **2016**, 1–14 (2016).

- Dziva, F., Christensen, H., Olsen, J. E. & Mohan, K. Random amplification of polymorphic DNA and phenotypic typing of Zimbabwean isolates of *Pasteurella multocida*. *Vet. Microbiol.* **82**, 361–372 (2001).
- Dziva, F., Muhairwa, A. P., Bisgaard, M. & Christensen, H. Diagnostic and typing options for investigating diseases associated with *Pasteurella multocida*. *Vet. Microbiol.* **128**, 1–22 (2008).
- Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
- Eisen, J. A. & Fraser, C. M. Phylogenomics: Intersection of evolution and genomics. *Science*. **300**, 1706–1707 (2003).
- Field, D. *et al.* Open software for biologists: from famine to feast. *Nat. Biotechnol.* **24**, 801–803 (2006).
- Friedberg, I., Kaplan, T. & Margalit, H. Evaluation of PSI-BLAST alignment accuracy in comparison to structural alignments. *Protein Sci.* **9**, 2278–2284 (2000).
- Galili, T. Data and text mining dendextend: an R package for visualizing , adjusting and comparing trees of hierarchical clustering. *Bioinformatics* **31**, 3718–3720 (2015).
- Galperin, M. Y., Makarova, K. S., Wolf, Y. I. & Koonin, E. V. Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res.* **43**, 261–269 (2015).
- García, N., Blocato, V., Repizo, G., Magni, C. & Lopez, P. in *Molecular Aspects of Lactic Acid Bacteria for Traditional and New Applications* 1–23 (2008).
- Giordano, A., Dincman, T., Clyburn, B. E., Steed, L. L. & Rockey, D. C. Clinical Features and Outcomes of *Pasteurella multocida* Infection. *Medicine (Baltimore)*. **94**, 1–7 (2015).
- Gunawardana, G. A., Townsend, K. M. & Frost, A. J. Molecular characterization of avian *Pasteurella multocida* isolates from Australia and Vietnam by REP-PCR and PFGE. *Vet. Microbiol.* **72**, 97–109 (2000).
- Guzmán, K., Rosadio, R., Maturrano, L. & Manchego, A. Asociación de Agentes Virales y Bacterianos en Cuadros de Neumonías. *Rev Inv Vet Perú* **24**, 524–536 (2013).
- Gyles, C. & Boerlin, P. Horizontally Transferred Genetic Elements and Their Role in Pathogenesis of Bacterial Disease. *Vet. Pathol.* **51**, 328–340 (2014).
- Hanage, W. P. & Gould, S. J. Not So Simple After All: Bacteria, Their Population Genetics, and Recombination. *Cold Spring Harb. Perspect. Biol.* **8**, 1–18 (2016).
- Harper, M., Boyce, J. D. & Adler, B. *Pasteurella multocida* pathogenesis: 125 Years after Pasteur. *FEMS Microbiol. Lett.* **265**, 1–10 (2006).
- Harper, M., Boyce, J. D. & Adler, B. in *Pasteurella multocida* 39–51 (2012). doi:10.1007/82-2012-202

- Harper, M. & Boyce, J. D. The Myriad Properties of *Pasteurella multocida* Lipopolysaccharide. *Toxins (Basel)*. **9**, 1–21 (2017).
- Harper, M. *et al.* A Heptosyltransferase Mutant of *Pasteurella multocida* Produces a Truncated Lipopolysaccharide Structure and Is Attenuated in Virulence. *Infect. Immun.* **72**, 3436–3443 (2004).
- Hatfaludi, T., Al-hasani, K., Boyce, J. D. & Adler, B. Outer membrane proteins of *Pasteurella multocida*. *Vet. Microbiol.* **144**, 1–17 (2010).
- Heddlestone, KL, Gallagher, JE, Rebers, P. Fowl Cholera: Gel Diffusion Precipitin Test for Serotyping *Pasteurella*. *Avian Dis.* **16**, 925–936 (1972).
- Horiguchi, Y. in *Pasteurella multocida* 113–129 (2012). doi:10.1007/82
- Hotchkiss, E. J., Hodgson, J. C., Leemput, E. S. De, Dagleish, M. P. & Zadoks, R. N. Molecular epidemiology of *Pasteurella multocida* in dairy and beef calves. *Vet. Microbiol.* **151**, 329–335 (2011).
- Hurtado, R. E. *et al.* Draft Genome Sequence of a Virulent Strain of *Pasteurella Multocida* Isolated From Alpaca. *J. Genomics* **5**, 68–70 (2017).
- Hurtado, R. Carhuaricra, D., Soares, S., Viana, M.V.C., Azevedo, V., Maturrano, L., Aburjaile, F. Pan-genomic approach shows insight of genetic divergence and pathogenic-adaptation of *Pasteurella multocida*. *Gene* **670**, 193–206 (2018).
- Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 1–11 (2010).
- Iraola, G. *et al.* Distinct *Campylobacter fetus* lineages adapted as livestock pathogens and human pathobionts in the intestinal microbiota. *Nat. Commun.* **8**, 1–8 (2017).
- Jabeen, S. *et al.* Complete Genome Sequence of *Pasteurella multocida* Serotype A Strain PMTB2.1 Isolated from Buffaloes That Died of Septicemia in Malaysia. *Genome Announc.* **5**, 1–2 (2017).
- Johnson, T. J. *et al.* Comparative genome analysis of an avirulent and two virulent strains of avian *Pasteurella multocida* reveals candidate genes involved in fitness and pathogenicity. *BMC Microbiol.* **13**, 1–10 (2013).
- Jolley, K. A. & Maiden, M. C. J. Using multilocus sequence typing to study bacterial variation: prospects in the genomic era. *Future Microbiol.* **9**, 623–630 (2014).
- Katoch, S., Sharma, M. & Patil, R. D. In vitro and in vivo pathogenicity studies of *Pasteurella multocida* strains harbouring different ompA. *Vet. Res. Commun.* **38**, 183–191 (2014).
- Kim, C., Song, S., Park, C. & Al, K. I. M. E. T. The D -Allose Operon of *Escherichia coli* K-12. *J. Bacteriol.* **179**, 7631–7637 (1997).
- Kim, M., Oh, H., Park, S. & Chun, J. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int. J. Syst. Evol. Microbiol.* **64**, 346–351 (2014).

- Kingston, A. W., Liao, X. & Helmann, J. D. Contributions of the  $\sigma$  W,  $\sigma$  M and  $\sigma$  X regulons to the lantibiotic resistome of *Bacillus subtilis*. *Mol. Microbiol.* **90**, 502–518 (2013).
- Kiu, R., Caim, S., Alexander, S., Pachori, P. & Hall, L. J. Probing Genomic Aspects of the Multi-Host Pathogen *Clostridium perfringens* Reveals Significant Pangenome Diversity, and a Diverse Array of Virulence Factors. *Front. Microbiol.* **8**, 1–18 (2017).
- Koehorst, J. J. *et al.* Comparison of 432 *Pseudomonas* strains through integration of genomic, functional, metabolic and expression data. *Sci. Rep.* **6**, 1–13 (2016).
- Koonin, E. V. Horizontal gene transfer: essentiality and evolvability in prokaryotes, and roles in evolutionary transitions. *F1000Research* **5**, 1–9 (2016).
- Kück, P. & Meusemann, K. FASconCAT: Convenient handling of data matrices. *Mol. Phylogenet. Evol.* **56**, 1115–1118 (2010).
- Kwaik, Y. A. & Bumann, D. Microreview Microbial quest for food in vivo: 'Nutritional virulence' as an emerging paradigm. *Cell. Microbiol.* **15**, 882–890 (2013).
- Lagesen, K. *et al.* RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* **35**, 3100–3108 (2007).
- Laikova, O. N., Mironov, A. A. & Gelfand, M. S. Computational analysis of the transcriptional regulation of pentose utilization systems in the gamma subdivision of Proteobacteria. *FEMS Microbiol. Lett.* **205**, 315–322 (2001).
- Laing, C. R., Whiteside, M. D. & Gannon, V. P. J. Pan-genome Analyses of the Species *Salmonella enterica*, and Identification of Genomic Markers Predictive for Species, Subspecies, and Serovar. *Front. Microbiol.* **8**, 1–16 (2017).
- Laslett, D. & Canback, B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* **32**, 11–16 (2004).
- Lee, C. W., Wilkie, I. W., Townsend, K. M. & Frost, A. J. The demonstration of *Pasteurella multocida* in the alimentary tract of chickens after experimental oral infection. *Vet. Microbiol.* **72**, 47–55 (2000).
- Li, L., Stoeckert, C. J. & Roos, D. S. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178–89 (2003).
- Liu, W. *et al.* Complete Genome Sequence of *Pasteurella multocida* HN06, a Toxigenic Strain of Serogroup D. *Genome Announc.* **194**, 3292–3293 (2012).
- Loman, N. J. & Pallen, M. J. Twenty years of bacterial genome sequencing. *Nat. Rev. Microbiol.* **13**, 1–9 (2015).
- López-Garrido, J., Puerta-Fernández, E., Cota, I. & Casadesús, J. Virulence Gene Regulation by L -Arabinose in *Salmonella enterica*. *Genetics* **200**, 807–819 (2015).
- Marchler-bauer, A. *et al.* CDD: NCBI's conserved domain database. *Nucleic Acids Res.* **43**, 222–226 (2015).



- Marttinen, P., Croucher, N. J., Gutmann, M. U., Corander, J. & Hanage, W. P. Recombination produces coherent bacterial species clusters in both core and accessory genomes. *Microb. Genomics* **1**, e000038 (2015).
- Marza, A. D. *et al.* The ability of lipopolysaccharide (LPS) of *Pasteurella multocida* B:2 to induce clinical and pathological lesions in the nervous system of buffalo calves following experimental inoculation. *Microb. Pathog.* **104**, 340–347 (2017).
- Massacci, F. R. *et al.* Characterization of *Pasteurella multocida* involved in rabbit infections. *Vet. Microbiol.* **213**, 66–72 (2018).
- Maximiliano, J. *et al.* Evaluación de la inmunogenicidad de una proteína recombinante de una *Pasteurella multocida* aislada de alpacas con neumonía. *Rev Inv Vet Perú* **29**, 339–348 (2018).
- May, B. J. *et al.* Complete genomic sequence of *Pasteurella multocida*, Pm70. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 3460–5 (2001).
- Mbuthia, P. G. *et al.* Time-course investigation of infection with a low virulent *Pasteurella multocida* strain in normal and immune-suppressed 12-week-old free-range chick. *Avian Pathol.* **40**, 629–637 (2011).
- McInerney, J. O., McNally, A., O'Connell, M. J., Lloyd, A. T. & Eyre-Walker, A. Why prokaryotes have pangenomes. *Nat. Microbiol.* **2**, 17040 (2017).
- McNally, A. *et al.* Combined Analysis of Variation in Core, Accessory and Regulatory Genome Regions Provides a Super-Resolution View into the Evolution of Bacterial Populations. *PLoS Genet.* **12**, e1006280 (2016).
- Mcvey, D. S. BRD research needs in the next 10 – 20 years. *Anim. Heal. Res. Rev.* **10**, 165–167 (2009).
- Michael, G. B. *et al.* ICE Pmu1, an integrative conjugative element (ICE) of *Pasteurella multocida*: structure and transfer. *J. Antimicrob. Chemother.* **67**, 91–100 (2012).
- Millán-aguiñaga, N. *et al.* Phylogenomic Insight into *Salinispora* (Bacteria, Actinobacteria) Species Designations. *Sci. Rep.* **7**, 1–11 (2017).
- Moraes, T. F. & Reithmeier, R. A. F. Membrane transport metabolons. *Biochem. Biophys. Acta* **1818**, 2687–2706 (2012).
- Morgan, J. K., Ortiz, J. A. & Riordan, J. T. The role for TolA in enterohemorrhagic *Escherichia coli* pathogenesis and virulence gene transcription. *Microb. Pathog.* **77**, 42–52 (2014).
- Moustafa, A. M. *et al.* Comparative Genomic Analysis of Asian Haemorrhagic Septicaemia-Associated Strains of *Pasteurella multocida* Identifies More than 90 Haemorrhagic Septicaemia-Specific Genes. *PLoS One* **10**, e0130296 (2015).
- Moye, Z. D., Zeng, L. & Burne, R. A. Fueling the caries process: carbohydrate metabolism and gene regulation by *Streptococcus mutans*. *J. Oral Microbiol.* **6**, 1–15 (2014).

- Mutters, R., IHM, P., Pohl, S., Frederiksen, W. & Mannheim, W. Reclassification of the Genus *Pasteurella* Trevisan 1887 on the Basis of Deoxyribonucleic Acid Homology, with Proposals for the New Species *Pasteurella dagmatis*, *Pasteurella canis*, *Pasteurella*. *Int. J. Syst. Bacteriol.* **35**, 309–322 (1985).
- Nawrocki, E. P., Kolbe, D. L. & Eddy, S. R. Infernal 1.0: inference of RNA alignments. *Bioinformatics* **25**, 1335–1337 (2009).
- Nolle, N., Felsi, A., Heermann, R. & Fuchs, T. M. Genetic Characterization of the Galactitol Utilization Pathway of *Salmonella enterica* Serovar Typhimurium. *J. Bacteriol.* **199**, 1–16 (2017).
- Oberhofer, T. R. Characteristics and Biotypes of *Pasteurella multocida* Isolated from Humans. *J. Clin. Microbiol.* **13**, 566–571 (1981).
- Okay, S. & Kurt Kızıldoğan, A. Comparative genome analysis of five *Pasteurella multocida* strains to decipher the diversification in pathogenicity and host specialization. *Gene* **567**, 58–72 (2015).
- Oksanen, J. *et al.* Package ‘vegan’. 1–295 (2018).
- Othman, S., Parton, R. & Coote, J. Interaction between mammalian cells and *Pasteurella multocida* B:2. Adherence, invasion and intracellular survival. *Microb. Pathog.* **52**, 353–358 (2012).
- Pacheco, A. R. *et al.* Fucose sensing regulates bacterial intestinal colonization. *Nature* **462**, 1–7 (2012).
- Paradis, E. *et al.* Package ‘ape’. 1–284 (2018).
- Passalacqua, K. D., Charbonneau, M. & Riordan, M. X. D. O. Bacterial Metabolism Shapes the Host – Pathogen Interface. *Microbiol. Spectr.* **4**, 1–21 (2016).
- Patané, J. S., Martins-JR, J. & Setubal, J. in *Comparative Genomics* 103–188 (2018).
- Peng, Z. *et al.* Genomic characterization of *Pasteurella multocida* HB01, a serotype A bovine isolate from China. *Gene* **581**, 85–93 (2016).
- Pessia, A., Grad, Y., Cobey, S., Puranen, J. S. & Corander, J. K-Pax2: Bayesian identification of cluster-defining amino acid positions in large sequence datasets. *Microb. Genomics* **1**, e000025 (2015).
- Petersen, A., Bisgaard, M., Townsend, K. & Christensen, H. MLST typing of *Pasteurella multocida* associated with haemorrhagic septicaemia and development of a real-time PCR specific for haemorrhagic septicaemia associated isolates. *Vet. Microbiol.* **170**, 335–341 (2014).
- Petersen, K. D., Christensen, H., Bisgaard, M. & Olsen, J. E. Genetic diversity of *Pasteurella multocida* fowl cholera isolates as demonstrated by ribotyping and 16S rRNA and partial atpD sequence comparisons. *Microbiology* **147**, 2739–2748 (2001).
- Petersen, T. N., Brunak, S., Heijne, G. Von & Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* **8**, 785–786 (2011).

- Polz, M. F., Alm, E. J. & Hanage, W. P. Horizontal gene transfer and the evolution of bacterial and archaeal population structure. *Trends Genet.* **29**, 170–175 (2013).
- Posada, D. jModelTest: Phylogenetic Model Averaging. *Mol. Biol. Evol.* **25**, 1253–1256 (2008).
- Prado, M. E., Dabo, S. M. & Confer, A. W. Immunogenicity of iron-regulated outer membrane proteins of *Pasteurella multocida* A:3 in cattle: molecular characterization of the immunodominant heme acquisition system receptor (HasR) protein. *Vet. Microbiol.* **105**, 269–280 (2005).
- Price, M. N., Dehal, P. S. & Arkin, A. P. Horizontal gene transfer and the evolution of transcriptional regulation in *Escherichia coli*. *Genome Biol.* **9**, 1–20 (2008).
- Rasko, D. A. *et al.* The Pangenome Structure of *Escherichia coli*: Comparative Genomic Analysis of *E. coli* Commensal and Pathogenic Isolates. *J. Bacteriol.* **190**, 6881–6893 (2008).
- Rímac, R., Luna, L., Hurtado, R., Rosadio, R. & Maturrano, L. Detection and genetic characterization of *Pasteurella multocida* from alpaca (*Vicugna pacos*) pneumonia cases. *Trop. Anim. Heal. Prod.* **49**, 1325–1328 (2017).
- Rodriguez-valera, F., Martin-Cuadrado, A.-B. & Lopez-Péres, M. Flexible genomic islands as drivers of genome evolution. *Curr. Opin. Microbiol.* **31**, 154–160 (2016).
- Rohmer, L., Hocquet, D. & Miller, S. I. Are pathogenic bacteria just looking for food? Metabolism and microbial pathogenesis. *Trends Microbiol.* **19**, 341–348 (2011).
- Rosadio, R., Cirilo, E., Manchego, A. & Rivera, H. Respiratory syncytial and parainfluenza type 3 viruses coexisting with *Pasteurella multocida* and *Mannheimia hemolytica* in acute pneumonias of neonatal alpacas. *Small Rumin. Res.* **97**, 110–116 (2011).
- Rosenbusch, C. & Merchant, I. A. A Study of the Hemorrhagic Septicemia Pasteurellae. *J. Bacteriol.* **37**, 69–89 (1938).
- Rouli, L., Merhej, V., Fournier, P. & Raoult, D. The bacterial pangenome as a new tool for analyzing pathogenic bacteria. *New Microbes New Infect.* **7**, 72–85 (2015).
- Roy, C. R. & Cherfils, J. Structure and function of Fic proteins. *Nat. Rev. Microbiol.* **13**, 631–640 (2015).
- Salusso, A. & Raimunda, D. Defining the Roles of the Cation Diffusion Facilitators in Fe<sup>2+</sup>/Zn<sup>2+</sup> Homeostasis and Establishment of Their Participation in Virulence in *Pseudomonas aeruginosa*. *Front. Cell. Infect. Microbiol.* **7**, 1–14 (2017).
- Sarangi, L. N. *et al.* Molecular Epidemiology of *Pasteurella multocida* Circulating in India by Multilocus Sequence Typing. *Transbound. Emerg. Dis.* **63**, 1–7 (2014).
- Schleif, R. AraC protein, regulation of the L-arabinose operon in *Escherichia coli*, and the light switch mechanism of AraC action. *FEMS Microbiol. Rev.* **34**, 779–796 (2010).

- Schliep, K., Potts, A., White, T. W., Stachniss, C. & Kendall, M. Package 'phangorn'. 1–81 (2018).
- Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
- Segerman, B. The genetic integrity of bacterial species: the core genome and the accessory genome, two different stories. *Front. Cell. Infect. Microbiol.* **2**, 1–8 (2012).
- Shakari-Garakani, A., Brinkkoter, A., Schmid, K., Turgut, S. & Lengeler, J. W. The genes and enzymes for the catabolism of galactitol, D-tagatose, and related carbohydrates in *Klebsiella oxytoca* M5a1 and other enteric bacteria display convergent evolution. *Mol. Gen. Genomics* **271**, 717–728 (2004).
- Shapiro, B. J. How clonal are bacteria over time? *Curr. Opin. Microbiol.* **31**, 116–123 (2016).
- Shivachandra, S. B., Viswas, K. N. & Kumar, A. A. A review of hemorrhagic septicemia in cattle and buffalo. *Anim. Heal. Res. Rev.* **12**, 67–82 (2011).
- Sivashankari, S. & Shanmughavel, P. Comparative Genomics -A perspective. *Bioinformation* **1**, 376–378 (2007).
- Sobczak, I. & Lolkema, J. S. The 2-Hydroxycarboxylate Transporter Family: Physiology, Structure, and Mechanism. *Microbiol. Mol. Biol. Rev.* **69**, 665–695 (2005).
- Soucy, S. M., Huang, J. & Gogarten, J. P. Horizontal gene transfer: building the web of life. *Nat. Rev. Genet.* **16**, 472–482 (2015).
- Stahl, M. *et al.* L-Fucose utilization provides *Campylobacter jejuni* with a competitive advantage. *Proc. Natl. Acad. Sci.* **108**, 7194–7199 (2011).
- Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
- Subaaharan, S., Blackall, L. L. & Blackall, P. J. Development of a multi-locus sequence typing scheme for avian isolates of *Pasteurella multocida*. *Vet. Microbiol.* **141**, 354–361 (2010).
- Tettelin, H. *et al.* Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial 'pan-genome'. *PNAS* **102**, 13950–13955 (2005).
- Tettelin, H., Riley, D., Cattuto, C. & Medini, D. Comparative genomics: the bacterial pan-genome. *Curr. Opin. Microbiol.* **11**, 472–477 (2008).
- Thibeaux, R. *et al.* Deciphering the unexplored *Leptospira* diversity from soils uncovers genomic evolution to virulence. *Microb. Genomics* **4**, 1–10 (2018).
- Townsend, K. M., Frost, A. J., Lee, C. W., Papadimitriou, J. M. & Dawkins, H. J. S. Development of PCR Assays for Species- and Type-Specific Identification of *Pasteurella multocida* Isolates. *J. Clin. Microbiol.* **36**, 1096–1100 (1998).

- Treangen, T. J., Ondov, B. D., Koren, S. & Phillippy, A. M. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol.* **15**, 1–15 (2014).
- Vernikos, G., Medini, D., Riley, D. R. & Tettelin, H. Ten years of pan-genome analyses. *Curr. Opin. Microbiol.* **23**, 148–154 (2015).
- Vos, M. & Didelot, X. A comparison of homologous recombination rates in bacteria and archaea. *ISME J.* **3**, 199–208 (2009).
- Vos, M., Hesselman, M. C., te Beek, T. A., van Passel, M. W. J. & Eyre-Walker, A. Rates of Lateral Gene Transfer in Prokaryotes: High but Why? *Trends Microbiol.* **23**, 598–605 (2015).
- Warnes, G. R. *et al.* Package ‘gplots’. 1–68 (2016).
- Watkins, E. R., Maiden, M. C. J. & Gupta, S. Metabolic competition as a driver of bacterial population structure. *Future Microbiol.* **11**, 1339–1357 (2016).
- Wickham, H., Chang, W. & RStudio. Package ‘ggplot2’. 1–189 (2016).
- Wiedenbeck, J. & Cohan, F. M. transfer and adaptation to new ecological niches. *FEMS Microbiol. Lett.* **35**, 957–976 (2011).
- Wilkie, I. W., Harper, M., Boyce, J. D. & Adler, B. in *Pasteurella multocida* **361**, 1–22 (2012).
- Wilson, B. A. & Ho, M. *Pasteurella multocida*: from zoonosis to cellular microbiology. *Clin. Microbiol. Rev.* **26**, 631–55 (2013).
- Yap, S. K., Zakaria, Z., Othman, S. & Omar, A. R. In vitro treatment of lipopolysaccharide increases invasion of *Pasteurella multocida* serotype B:2 into bovine aortic endothelial cells. *J. Vet. Sci.* **19**, 207–215 (2018).
- Yu, C. *et al.* Genome sequencing of a virulent avian *Pasteurella multocida* strain GX-Pm reveals the candidate genes involved in the pathogenesis. *Res. Vet. Sci.* **105**, 23–27 (2016).
- Yue, M. *et al.* Allelic variation contributes to bacterial host specificity. *Nat. Commun.* **6**, 1–11 (2015).

## 10. GLOSARIO

**Familia de genes ortólogos:** Se refiere a grupo de genes ortólogos que son identificadas de múltiples genomas como se ilustra en la figura 1.

**Pangenoma:** La colección completa de familias de genes ortólogos que son encontradas en una especie dada.

**Genoma core:** La colección de familias génicas que son encontrados en todos los miembros de una especie en particular.

**Genoma accesorio:** La colección de familias génicas que son encontrados en algunos, pero no en todos los genomas de las especies.

**Análisis pangenómico:** Uso de un conjunto de herramientas bioinformáticas para estudiar el pangenoma de un organismo.

**Filogenómica:** Reconstrucción de la historia evolutiva de organismos tomando en cuenta todo el genoma. Para el presente, un “genoma” es entendido como su contenido genético (genes codificantes de proteínas).